AFIT/GLM/LSR/92S-22

A COMPARISON OF COMPUTER AND PAPER MEDIA IN PSYCHOLOGICAL TESTING

THESIS

Anthony C. Hensley, Captain, USAF
Timothy R. Morris, Captain, USAF

AFIT/GLM/LSR/92S-22

Approved for public release; distribution unlimited

The views expressed in this thesis are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

AFIT/GLM/LSR/92S-22

A COMPARISON OF COMPUTER AND PAPER MEDIA IN PSYCHOLOGICAL TESTING

THESIS

Presented to the Faculty of the School of Systems and Logistics

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Logistics Management

Anthony C. Hensley, B.S.

Captain, USAF

Timothy R. Morris, B.A.

Captain, USAF

September 1992

Preface

The purpose of this study was to examine differences based on mode of administration in subject responses to psychological tests. Specifically, the purpose was to examine if there are significant differences in the way people respond to computer and paper versions of psychological tests. The research is important to the Air Force and the Department of Defense due to the massive implementation of computer technologies in all aspects of military operations. It is essential that the implementation of new technologies be carefully examined to determine if the technology is meeting the requirements it is designed for and to determine what impact the new technology will have on the workforce which will utilize it.

The idea for this research was based on a suggestion by our thesis advisors, Major Wayne Stone and Dr. Guy Shane. Our thanks to Major Stone for the initial idea for this research and his continual guidance throughout the project. Also thanks to Dr. Shane for his patience and expert assistance in carrying out the research effort.

Special thanks to Dr. Theodore L. Hayes, Assistant Professor, Department of Psychology, Wright State University (WSU), Ohio. Dr. Hayes provided guidance and invaluable support to our research at WSU. Through Dr. Hayes' efforts and coordination we were able to arrange for research approval, subjects, and computer facilities for data collection at WSU.

This research would not have been possible without the support of E. F. Wonderlic and Associates and Psychological Assessment Resources. These companies provided us the use of their instruments and all related materials in carrying out our research.

At E. F. Wonderlic and Associates we would like to acknowledge the support and assistance of Mr. Fred Rafilson, Mr. Eric Schnell, and Mr. Victor Artese. Mr. Rafilson provided initial guidance, support and approval for the project. Thanks to Mr. Schnell for all the correspondence and patience in dealing with our hardware and software compatibility problems. Mr. Artese solved many of the technical problems we faced as a result of the compatibility difficulties. Special thanks to E. F. Wonderlic and Associates for implementing the programming change necessary to meet our requirements.

At Psychological Assessment Resources we would like to acknowledge Dr. Bob Smith, III, Dr. Glenn Curtis, Ms. Brenda Mercer, and Ms. Karen Tuttle. Thanks to Dr. Smith and Dr. Curtis for their approval and support of the research. Thanks to Ms. Mercer for the prompt and attentive support in responding to our requests; thanks to Ms. Tuttle for her patience in assisting us with our many technical problems.

And finally, we would like to thank our families for their understanding and support. Captain Morris would like to thank his wife, Tracy, and daughter Emily who unselfishly supported him throughout the entire fifteen months of AFIT. Without their love and support this research would not have been possible.

Captain Hensley would like to thank his wife, Sharon, and daughter Emily who understood and supported him through the long days and weekends spent studying. Thanks also to Ann, John, and Noah for being there when needed.

Anthony C. Hensley and Timothy R. Morris

Table of Contents

Page

List of Tables

Abstract

The purpose of this study was to examine differences based on computer and paper modes of administration in subject responses to psychological tests. The hypotheses tested were the expectations that subject's scores on the computer administration of the Wonderlic Personnel Test (WPT) would be higher than the paper-and-pencil administration and that the subjects would report higher levels on the various scales of the Vocational Preference Inventory (VPI). A counterbalance experimental design with two experimental groups was used. The subjects were all volunteers from the Air Force Institute of Technology (AFIT) and Wright State University (WSU). A total of 90 subjects completed the study. This included 75 students from AFIT and 15 from WSU. 69 subjects were male and 21 were female. Subjects from each institution were randomly assigned to two experimental groups. The administration mode of the psychological tests was the independent treatment condition for the study.

Based on the analysis and findings, it was concluded that:

There is no statistically significant difference between the modes of administration, paper-and-pencil and computer, for either the Wonderlic Personnel Test or the Vocational Preference Inventory.

There was a significant order effect with the response variable WPT. This significance can be attributed to the fact that the WPT is a cognitive test and there was a practice effect. The subjects scored significantly higher on the second administration of the WPT regardless of the mode of administration.

There were also significant differences in four of the VPI response variables between the experimental groups. The four response variables were the VPI Investigative, Artistic, Infrequency, and Acquiescence scales. These findings can be attributed to the nature of the VPI instrument. Because the VPI is a self-reported inventory, the differences between the experimental groups may be attributed to the composition of the experimental groups.

Subjects reported that they preferred the computer administration mode over the paper-and-pencil administration mode 54% to 17%. Additionally, 50% of the subjects reported that the computer versions were "easier to complete", while 13% reported the paper-and-pencil versions were "easier to complete".

A COMPARISON OF COMPUTER AND PAPER MEDIA IN PSYCHOLOGICAL TESTING

I. Introduction

General Issue

In recent years, the use of the computer as an instrument of data collection, assimilation, and analysis has dramatically increased. The rapid increases in computer technology have made computer applications more accessible. "Increases in computer availability, speed, and data storage and retrieval capabilities and decreases in cost have been dramatic . . . ." (Allred, 1986:6) Many fields have realized the potential that computer applications can hold. The field of psychology is among those fields. In fact, "the application of computers to interviewing, testing, and interpretation is so widespread that it is noteworthy if computers are not used" (Waller and Reise, 1989:1051). With this growth in the applications of computer technology, an examination of the significance and ramifications of computer use for test administration is required to confirm or deny the equivalence of computer-aided testing to the more traditional paper-and-pencil administration.

The potential resulting effects of research of this type in the field of psychology are unlimited. New methods of adapting the computer to various psychological measurements will be discovered. Furthermore, research on the interaction of computer

and user from the field of psychology carries implications important to the study of the overall effects of rapid advancements made in computer technologies.

The significance and effects of the rapid growth of computer technologies is important to the Air Force as well. One need only peer into any office environment to discover the extent of computer usage in today's Air Force. Complete automation in the workplace appears to be on the horizon. It is important, then, to understand the potential effects of widespread computer technologies on the workforce. The study of human interaction with traditional and computer technologies is a vital precursor toward forming an understanding of the onset of automation on the workforce.

Specific Problem

This research focuses on the use of computers as instruments of data collection in psychological testing. Specifically, based on the results of the analysis, the correlation between traditional, paper-and-pencil applications and computer applications as a method of response collection for the purpose of psychological measurement were examined.

Investigative Questions

(1) Is there a significant correlation between subject responses when a traditional, paper-and-pencil instrument is used compared with a computer instrument of the same psychological test?

(2) Do subjects respond differently to computer instrument stimuli compared with traditional instrument stimuli, affecting the nature of subject response to the psychological measurement?

(3) Does the subject's level of computer experience affect the nature of the subject's response to the computer instrument compared with the traditional instrument?

Scope of the Research

While the application of computer technologies is extremely broad, this paper is limited to the discussion of computers in psychological testing. This research provides an examination of the difference in subject responses which are realized as a result of the difference in test administration modes. Specifically, the differences in the administration modes of the Wonderlic Personnel Test (Wonderlic, 1983) and the Vocational Preference Inventory (Holland, 1985) were examined. This study was not based on the interpretation of the individual scores of the subjects tested. Instead, individual test scores were aggregated in order to perform statistical analysis on the means of each test measure and the administration modes. In this regard, the score of each individual's test is important only as it is a member of the aggregate group.

Thesis Organization

Following this introduction, Chapters II, III, and IV provide a review of the relevant literature, details of the research method, and a presentation of the results and findings of the analyses, respectively. Included in Chapter IV are statistical tables summarizing the results. Chapter V presents interpretations and conclusions from the experiment results, and makes recommendations for future research.

II. Literature Review

This chapter examines the relevant literature related to the use of computers in psychological testing. The review consists of documentation published in English within the past ten years. A search of the *Social Sciences Index* as well as the Dialog and NASA/Recon databases led to the references presented herein. In reviewing the literature, the following subject areas were addressed: (1) Comparison of Computer and Traditional Applications, (2) Ethics and Standards in Computer Administration, and (3) Present and Future Applications of Computer-Based Tests.

Discussion of the Literature

Comparison of Computer and Traditional Applications. The majority of the literature reviewed expressed confidence in the ability of computer applications to measure psychological effects. A number of the authors conducted experiments to determine whether the correlation between the two applications was significant. The findings revealed that there is typically a high correlation between the results found with the two methods and that computer applications often yield equal or higher reliability (Bagley and Genuis, 1991:287; Brown, 1984:455; Elithorn, Mornington, and Stavrou, 1982:247; Kennedy, Baltzley, Turnage, and Jones, 1989a:1067; Kennedy, Baltzley, Wilkes, and Kuntz, 1989b:1270; McCaughan, 1982:92-100; Thompson and Wilson, 1982:280; Vale, 1981:405; Waller and Reise, 1989:1056; Wilson, Thompson, and Wylie, 1982:294). Brown noted that "administration of test instruments by computer was substantially more reliable than traditional pencil and paper techniques" (1984:455).

Waller and Reise's results suggest that "computerized adaptive personality assessment works impressively well" (1989:1051).

Bagley and Genuis examined a computerized questionnaire that investigates the prevalence of child sexual abuse. They found that the computer questionnaire "elicited significantly more recall of prior abuse than a paper questionnaire" (1991:287). This study was based on subject pools of studies conducted in 1984 and 1990, respectively. While their finding show a significant difference in subject responses based on the administration modes, they recommend further research utilizing "counterbalanced order of administrations to the same subjects" (1991:288). Additionally, they report that "91% said they would prefer the computerized questionnaire to a self-completion method or a personal interview" (1991:287).

Kennedy *et al.* selected 11 tests from two microcomputer-based performance test batteries and evaluated them against the Wonderlic Personnel Test. The Wonderlic Personnel Test is a "timed paper-and-pencil test of the ability to learn" and "is often considered an abbreviated index of global intelligence" (1989a:1059,1063). They found that:

> The ability of this short performance battery to identify 'factors of the mind' with consistency over time and to predict, validly, criterion performance on a short measure of intelligence attest to the greater utility of batteries which have been developed specifically to capitalize on the necessary psychometric properties of reliability and stability. (1989a:1072)

While the majority of the authors expressed confidence in the significance of the correlation between computer and traditional applications, a number of the authors did have validity questions. Furthermore, some of the authors expressed concern over the use of normative data derived from traditional tests in the interpretation of computer-based

tests (Allred, 1986:38-48; Brown, 1984:456; Burke and Normand, 1987:46-47; Guastello, Guastello, and Craft, 1989:477-478; Lautenschlager and Flaherty, 1990:312-314; Matarazzo, 1986:14,21).

Allred notes that computerized psychological testing is now commonplace. However, there is "little or no evidence that computerized tests are equivalent to their paper-and-pencil form . . ." (1986:1782). Allred researched this hypothesis by examining a commercially-available computerized form of Gough's Adjective Check List (ACL); ACL is a personality test. She found that subjects checked significantly "more adjectives as self-descriptive, on the average, than they checked on the conventional paper-and-pencil form of the ACL" (1986:119). Allred concluded that this difference would invalidate the use of existing norms (1986:119-124).

Brown also cited the potential problem of the "use of existing paper-and-pencil testing techniques on computers without the generation of normative data based specifically on these computer applications" (1984:456).

Burke and Normand address this concern by stating that "if the computer administration of a test is equivalent to the conventional administration, then norms developed with the conventional test can be used to interpret scores obtained by computers" (1987:47).

Clearly, a number of authors have expressed concerns that normative data collected for the paper-and-pencil versions of psychological tests are being used for the computer versions of the instruments. Further examination of the equivalence of the modes of administration is necessary before the norms associated with paper-and-pencil tests can be used for computer versions of the tests.

Ethics and Standards in Computer Administration. While the majority of the authors agree that there is a significant correlation between traditional and automated administration of tests, a number of the authors express concern over the use of computerized findings (Burke and Normand, 1987:48; Farrell, 1989:1-11; Gaustello *et al.*, 1989:477-478; Hedlund, 1987:32-33; Hedlund, 1988:23; Kramer, 1987:889; Matarazzo, 1986:14,18-23; Space, 1981:599-600; Walker and Myrick, 1985:51-56). Walker and Myrick state that, "the issue is not whether computers *should* be used to score and interpret psychological tests but under what conditions and guidelines are such procedures to be used" (1985:51).

Matarazzo calls for the establishment of standards and guidelines to govern the use of computerized psychological tests. Specifically, standards and guidelines are needed in the use of computerized test interpretations. Matarazzo states that "automated and 'canned' psychological interpretations do not offer plausible interpretations from identified clinicians based on the same test findings". He contends that the interpretations should be based on "identified" clinicians' interpretations of duplicate test findings. (1986:20)

Burke and Normand also feel that the use of computer interpretations is problematic. "Only providing a narrative report for a clinically based test without the option of having a professional assist in interpreting the test results is under no circumstances sufficient and proper feedback" (1987:48). Farrell echoes these concerns and adds that "existing guidelines for evaluating computer applications within psychology are not having a sufficient impact on professional practice" (1989:1).

While Kramer does feel that "clearly CBTI (computer-based test interpretations) present questions and challenges not before encountered . . . this does not imply, however,

that we should proceed lock step with the development of guidelines . . ." (1987:889). Kramer feels that the existing standards and guidelines can be revised to cover this emerging technology and that "we should not saddle CBTI with separate guidelines simply because it is the application of a powerful technology" (1987:890).

Present and Future Applications of Computer-Based Tests. Clearly, none of the authors express the opinion that computerized psychological tests should be discontinued or banned. Additionally, the majority of the authors are quite excited about the possibilities computer applications can add to psychological testing, although some had reservations.

In the study by Bagley and Genuis, mentioned previously, they concluded that a strong case has been made for the "increased use of computers for assessment in psychiatry and psychology; many methodological issues have now been overcome," and their findings suggest that the computer can be a potentially powerful tool with "young, literate populations" (1991:288).

Brown notes that computerized tests yield:

> . . . positive benefits including (a) better client response to the testing situation, (b) cost effectiveness, (c) the ability of the computer to do interactive testing, (d) the generation of standardization data, (e) more efficient use of staff time, (f) more efficient and accurate scoring, (g) reduced error rates in scoring and administration, (h) validity of interpretation of results, and (i) potential assistance to persons with visual and/or auditory handicaps. (1984:455-456)

Burke and Normand echo the potential of computerized tests "being practical, cost-effective, and psychometrically sound means of assessing individuals" (1987:49). However, they state that this potential can only be realized "if proper considerations are

made in designing, developing, and implementing these testing systems and if professional standards are maintained by computer test service providers and users" (1987:49).

Elithorn *et al.* contend that "in computerizing psychological test systems one should aim to achieve more than mere automation". They state that tests should not be automated for the sole purpose of automating, but rather, based on the "merits of the test and their suitability for automation" (1982:247,249).

Farrell notes that applications "which capitalize on the full power of the computer have only begun to be developed" (1989:10). He believes that new innovations in the application of computer power will require guidelines and standards to enforce professional practices. He warns that practicing psychologists should be "sensitive to the issues raised by these new developments". Furthermore, "guidelines will be needed as these innovations are introduced into professional practice" (1989:1,10).

These concerns regarding guidelines, normative data, and utilization of computer capabilities are echoed by a number of authors; however, the authors are generally optimistic about the future of computerized psychological testing. (Hedlund, 1988:23; Kennedy *et al.*, 1989b:1070-1072; Kramer, 1987:889; Matarazzo, 1986:14,23; McCaughan, 1982:101-106; Space, 1981:602-604; Thompson and Wilson, 1982:285-288; Vale, 1981:399; Volans, 1982:301-304; Walker and Myrick, 1985:51,56; Waller and Reise, 1989:1051,1056-1057; Wilson *et al.*, 1982:294-296).

A number of the authors suggest that the appropriate means of creating computerized tests is not to computerize existing tests, but to specifically design tests for the computer (Space, 1981:603; Thompson and Wilson, 1982:287; Volans, 1982:301-304). Space notes that "to use the computer effectively, tests designed especially for [the

computer] will add depth to any test battery" (1981:603). Furthermore, Space adds that computerizing existing tests may have normative problems, may not use the computer's capabilities, and may not be cost-effective (1981:603). Thompson and Wilson add that "the real potential lies in making computers do those things they do well and this will increasingly include new tests which exploit the computer's capacity to analyze data quickly" (1982:287).

Another advantage of computer designed and implemented tests is the ability to perform adaptive testing. Vale defines adaptive testing as follows:

> Adaptive, or tailored, testing is a relatively new form of psychological testing in which a test is tailored to an individual during the testing process, such that those items that are most appropriate to the individual and most informative about the characteristic being measured are administered. (1981:399)

Volans discusses the advantages of using interactive automated testing systems to conduct tailored testing. She notes that tailored tests are "likely to prove more useful, and less stressful and time-consuming than conventional tests in applied and clinical settings" (1982:302). She further notes that this approach "poses a number of challenges at the design stage" (1982:302).

Thompson and Wilson note that designing computer adaptive tests will allow for the generation of "stimuli of known complexity . . . and alternate forms of tests". The generation of alternative (parallel) forms can help "overcome retest effects". The retest effect comes from changes in the results of the test due to the subject learning the test through repetition. Using the computer's storage capability, "carefully controlled banks of material" could be used to generate alternative forms. They note that sensitive coding

means are available which can easily be used "to produce multiple forms suitable for repeated testing". (1982:286)

These findings support the need for development of new tests, as opposed to computerizing existing tests, because these test complexities are best suited for computer application (Volans, 1982:303).

Conclusions

It is evident from the current literature that computer-based psychological tests can and will provide new and powerful applications in the profession. While there have been some studies into the correlation of findings between traditional, paper-and-pencil applications and computerized applications, these studies have been limited. Further examination of the differences between paper-and-pencil and computer applications is warranted. The literature clearly reveals that computer-based psychological testing offers many potential benefits to the profession.

With these potential benefits come a number of concerns. An example is the concern that these computer-based programs are being used with little or no research into their validity. Furthermore, the normative data of the traditional application is being used for interpretations of the computer-based results. The questions of validity and norms warrant further examination.

There is also an opinion that the power of the computer is not being used to the extent possible (Space, 1981:603; Thompson and Wilson, 1982:287; Volans, 1982:301-304). Specifically, the majority of the computer-based tests are simply computerized versions of traditional tests. These authors feel that computer-based tests should be especially designed to take advantage of the capabilities of the technology. Simple

"computerization" of existing tests fails to use the capabilities of the computer and may not be cost-effective. Space noted this with an examination of a computerized version of the Minnesota Multiphasic Personality Inventory (MMPI), where a computer terminal was tied-up for 1.5-2 hours "answering sequential true-false questions, when an optical scanning of pencil answers could do as well" (Space, 1981:603). It should be noted that Space's research was conducted in 1981 when computer resources were not as readily available or as cost-effective as they are presently. However, the underlying trade-off decision remains the same; does the benefit of the computerized format of the test warrant the cost of the computer resources required for its administration?

Adaptive testing is one area where the computer can provide significant advances. The adaptive tests are tailored during the test to allow the examination of the most important and informative measures. Using adaptive testing, the number of questions and question selection are based on the subject's response to previous questions. Conversely, traditional tests are based on a fixed number and sequence of questions. The speed and storage capacity of the computer allow for this type of testing to occur. Adaptive testing can potentially provide more insight into the actual measurement which is under examination.

Another benefit of computerized testing is the ability of the computer to randomize the items presented. This method allows parallel forms of the test to be administered from a database of items. The benefit is that a subject can be retested without suffering retest effects. With both adaptive testing and parallel form databases, examination of the application of existing normative data and the instrument's reliabilities are necessary. The instrument's reliabilities will be difficult to estimate based on existing normative data.

To estimate the reliabilities, every possible combination of items must be examined. With parallel form databases this would not be as difficult as with adaptive testing. In adaptive testing the possible number of item combinations could rapidly approach unmanageable numbers. While adaptive testing seems to truly exploit the capabilities that the computer technology offers, from a psychometric standpoint, the development and acceptance of the instrument will be problematic due to the difficulties in estimating reliability and developing normative data.

Clearly, there is a future for computer-based tests within the field of psychology. However, as with the application of any new and powerful technology, careful attention must be paid to its application. Further research and oversight must take place within the profession to ensure that the technology is applied responsibly.

Finally, additional study into the correspondence of paper-and-pencil to computer-based tests, the validity of these tests, and the use of existing normative data must occur.

III. Method

This chapter details the method used to examine the investigative questions presented in Chapter 1. The experiment was conducted at two separate locations; computer classrooms at the Air Force Institute of Technology and the Educational Resources Center at Wright State University were utilized. A quasi-experimental design (Campbell and Stanley, 1963:34) was used with subjects randomly assigned to two experimental groups. The specific details concerning the subjects, materials, procedures, and statistical analyses are presented below.

Subjects

The subjects for this study consisted entirely of volunteers. These volunteers were students enrolled in introductory psychology courses at Wright State University (WSU) and graduate students from the School of Systems and Logistics, Air Force Institute of Technology (AFIT). Subjects from each of the two institutions were randomly assigned to the experimental groups.

There were initially 127 total volunteers with 40 from WSU and 87 from AFIT. 104 subjects completed the first phase of the study and 90 completed all phases of the research. Of the 90 total subjects 15 were from WSU and 75 were from AFIT. The final subject population consisted of 21 females and 69 males. The breakdown of each group by institution and gender is reported in Table 1 and subject ages are reported in Table 2. Based on the subjects' reported educational background, 83% of the subjects had bachelor's degrees or higher and 17% were undergraduate students.

TABLE 1

SUBJECT FREQUENCY BY GROUP, INSTITUTION, AND GENDER

Subject Population - AFIT

Gender	Group 1	Group 2	Total
Male	34	32	66
Female	5	4	9
Total	39	36	75

Subject Population - WSU

Gender	Group 1	Group 2	Total
Male	1	2	3
Female	10	2	12
Total	11	4	15

While all subjects participated in the experiment on a voluntary basis, the subjects from WSU received 4 research credits as an incentive for their participation. Incentive credits were granted based on the standard WSU participation incentive policy of one credit per thirty minutes of student participation in department-sponsored studies. Subjects from AFIT received no incentive for their participation in the study.

TABLE 2

SUBJECT FREQUENCY BY AGE AND GROUP

Age	Group 1	Group 2	Total
20 or less	8	3	11
21 - 25	1	1	2
26 - 30	18	12	30
31 - 35	13	17	30
36 - 40	8	5	13
41 - 45	2	2	4
Total	50	40	90

Materials

The materials used in this experiment include two separate psychological tests which were administered in both paper-and-pencil and computerized formats. These tests are the Wonderlic Personnel Test and the Vocational Preference Inventory. In addition to these tests, a paper-and-pencil pre-experiment questionnaire and a paper-and-pencil post-experiment questionnaire were used in the collection of data.

Wonderlic Personnel Test. Test 1 was the Wonderlic Personnel Test (WPT), Form IV. This instrument is a twelve-minute timed test and contains 50 total items. The WPT is a general mental ability test which was "designed and created for testing adults in business and industrial situations" (Wonderlic, 1983:3). The test consists of quantitative,

verbal, and spatial questions. The subjects are instructed to record their response to each question in a space provided. Some questions require multiple choice-type answers, while others question require the subject to perform a calculation and provide the written answer directly.

The test booklet for the paper-and-pencil version of this instrument contains both the questions and the answer sheet. Subjects are required to write their responses on a space provided to the right of each question. This version is scored using a scoring key provided with the instrument.

The computerized version of the WPT requires the subject to input their response using a computer keyboard. This version also permits the subject to review their responses, skip questions, and return to questions as they desire. Additionally, the computerized version has a tutorial, which was utilized in this experiment, to educate the subject in the operation of the program. This version is self-scoring with the computer program calculating the subject's score. The researcher is able to examine the actual subject response to each item by examining an audit/validity file generated by the software.

The test was administered in accordance with the standardized procedures specified in *Wonderlic Personnel Test Manual*, (Wonderlic, 1983:4). Both versions of the WPT yield a raw score based on the total number of correct responses. This raw score was the measure of interest (dependent variable) for this research.

WPT Reliabilities. The *Wonderlic Personnel Test Manual*, (Wonderlic, 1983:8) reports the test's reliabilities. The manual specifies test-retest reliabilities, longitudinal reliabilities, alternate forms reliabilities, and internal consistency. The test-

retest reliabilities and internal consistency are of particular interest to this research. The alternate forms reliability is not relevant because the same form of the test was used in all administrations. The longitudinal reliability is not relevant due to the short time span of this research. Test-retest reliability provides a measure of the question, "if a person is tested again shortly after the first test will the second score be sufficiently the same as the first to yield the same interpretation?" (Wonderlic, 1983:8) The internal consistency provides a measure of the question, "if two people score the same, what is the likelihood that they may have answered different questions and therefore each score has a different interpretation?" (Wonderlic, 1983:8) The manual reports the test-retest reliabilities to range between .82 and .94 and the internal consistency measures to range from .88 to .94.

This research examined the internal consistency of both the paper-and-pencil version and the computer version of the WPT utilizing the split-half technique. Emory and Cooper note that "the split-half technique can be used when the measuring tool has many similar questions or statements to which the subject can respond" (1991:187). The split-half technique involves analysis in which "the results are separated by item into even and odd numbers or into two randomly selected halves" (1991:187).

For both administration modes, the results were separated into two halves by even and odd numbered items. A correlation of the two halves was performed utilizing the Spearman-Brown correction formula "to adjust for the effect of test length and to estimate the reliability of the whole test" (Emory and Cooper, 1991:187).

The Spearman-Brown correction formula can be found in most psychometric textbooks, for example, Hopkins and Stanley (1981:126). Applying the Spearman-Brown correction formula, the internal consistency coefficient computation for both the paper-

and-pencil and the computer versions of the WPT was found to be .89. These coefficients are within the .88 to .94 range as reported in the *Wonderlic Test Manual.*

The test-retest coefficient for the WPT is controlled for mode of administration due to the counterbalanced experimental design. The retest coefficient was calculated to be .83. This value is also consistent with the .82 to .94 range specified in the test manual.

Vocational Preference Inventory. Test 2 is the Vocational Preference Inventory (VPI) - 1985 Revision, which contains 160 total items comprising eleven scales. The 160 items are a listing of occupations. The subject responds to each item by noting if they find the occupation interesting or appealing or if they find the occupation uninteresting or unappealing. Additionally, the subject may leave an item blank if indifferent to that item.

The paper-and-pencil version of this instrument provides a test booklet which lists the 160 items. The subjects are instructed to record their preferences on a separate answer sheet. This answer sheet contains circles at each item containing "Y" and "N" responses which the subject darkens appropriately with a pencil. This instrument is scored manually using a scoring key which is provided.

The computerized version of the instrument presents the items on the computer screen and the subjects are instructed to press the "Y" key for items that interest them and the "N" key for items that don't interest them. Subjects are instructed to press the "SPACEBAR" for items to which they are indifferent. The computer automatically scores the test and provides the score for each scale. The version of the software used in this

experiment did not provide the researchers with the capability to examine the subject's response to each individual item.

The test was administered in accordance with the standardized procedures specified in *The Vocational Preference Inventory: Professional Manual*, (Holland, 1985:2-3). The VPI yields eleven scales which were used as the measures of interest (dependent variables) for this research. The eleven VPI scales are: Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E), Conventional (C), Self-Control (Sc), Masculinity-Femininity (Mf), Status (St), Infrequency (Inf), and Acquiescence (Ac). Appendix C presents the clinical interpretations of each of the scales as reported in the test manual.

VPI Reliabilities. *The Vocational Preference Inventory: Professional Manual*, (Holland, 1985) reports the test's reliabilities. The internal consistency (KR 20) of each of the VPI scales is provided in Table 3. The coefficients for each scale indicate that "the content of most scales is relatively homogenous". The exceptions are the "Masculinity, Status, and Infrequency scales which are composed of relatively heterogeneous occupations" (Holland, 1985:3).

The test-retest reliabilities of the scales are provided in Table 4. One study reported in the test manual indicates that the "retest reliabilities for the interest scales range from .54 to .80, with a median of .71, for a sample of junior college students (62 men and 53 women), for a 3 month interval" (Holland, 1985:3).

TABLE 3

INTERNAL CONSISTENCY COEFFICIENTS FOR VPI SCALES (HOLLAND, 1985:3)

	Men		Women	
Scale	**KR 20**	***n***	**KR 20**	***n***
Realistic	.85	171	.87	187
Intellectual	.91	176	.91	158
Artistic	.90	183	.90	165
Social	.90	186	.88	167
Enterprising	.87	190	.86	165
Conventional	.88	192	.90	161
Self-Control	.85	186	.81	172
Masculinity	.56	175	.53	159
Status	.62	174	.58	162
Infrequency	.42	178	.55	154

TABLE 4

RELIABILITY COEFFICIENTS (RETEST) FOR VPI SCALES (HOLLAND, 1985:3)

	Adult Women (Mean Age = 40.7)		College Seniors	College Freshmen
	(*n* = 31)	(*n* = 28)	(*n* = 17)	(*n* = 26)
Scale	**2 wks.**	**2 mons.**	**6 wks.**	**1 yr.**
Realistic	.79	.57	.92	.86
Intellectual	.71	.71	.83	.65
Artistic	.73	.69	.98	.73
Social	.72	.66	.79	.76
Enterprising	.65	.79	.78	.71
Conventional	.83	.84	.74	.61
Self-Control	.79	.58	.86	.84
Masculinity	.75	.78	.85	.82
Status	.72	.70	.62	.84
Infrequency	.66	.80	--	.78
Acquiescence	.71	.62	--	.93

The reliability coefficients, based on this research, for the VPI scales are summarized in Table 5; reported are the internal consistency coefficients which were calculated utilizing the split-half technique. The internal consistency coefficients are reported for the paper-and-pencil mode only due to the inability to collect the individual item responses on the computerized version of the VPI.

The test-retest coefficients are based on a correlation of Administration 1 scale scores with Administration 2 scale scores. Due to the counterbalanced experimental design, the test-retest coefficients are controlled for administration mode. The coefficients are comparable to coefficients reported in Holland (1985).

TABLE 5

RELIABILITY COEFFICIENTS FOR THE VPI SCALES

	Internal Consistency		Test-Retest (n = 90)
Scale	r'_{xx}	n	r
Realistic	.91	79	.80
Intellectual	.89	77	.85
Artistic	.81	79	.76
Social	.86	79	.82
Enterprising	.87	76	.78
Conventional	.85	75	.85
Self-Control	.90	75	.82
Masculinity	.60	96	.81
Status	.63	96	.76
Infrequency	.69	96	.75
Acquiescence	.74	78	.77

Questionnaires. An 18-item pre-experiment questionnaire, adapted from Casey and Kveene (1991), and a six-item post-experiment questionnaire, developed by the authors, were administered to all subjects.

The pre-experiment questionnaire was designed to collect demographic data and to assess the subject's reported level of computer experience. The demographic information collected included the subject's name, sex, age group, and eduction level. Subjects responded to the following age groupings: 20 or less, 21-25, 26-30, 31-35, 36-40, 41-45, and 45 or more.

Pre-experiment questionnaire items 11-18 were used to assess the subject's computer experience levels. Subjects responded to these items utilizing a scale which contained six experience levels. Scale item 1 and scale item 2 were rearranged to correct for an ordering error in the scale. Subject's responses to questionnaire item 11-18 were summed to establish an independent variable for the study. Internal consistency of this measure was estimated utilizing Cronbach's alpha; the alpha for these items was .91.

The post-experiment questionnaire was designed to report the subjects' preferences for the two administration modes.

Procedure

Each test was administered to all subjects in both paper-and-pencil instrument and computer instrument formats. The questionnaires were administered to all subjects in paper-and-pencil format only.

Experimental Design. The study consisted of a counterbalanced experimental design which was controlled for time. This design involved two experimental groups which completed opposite modes (computer and paper) of the tests during each of two

administrations. As shown in Table 6, "Counterbalanced Experimental Design", during Administration 1, Group 1 completed the computer version of the WPT and Group 2 completed the paper version of the WPT. During Administration 2, both groups completed the respective version not completed during Administration 1. This is the counterbalance; Group 1 takes the WPT-computer during A_1 and Group 2 takes the WPT-computer during A_2. To examine the differences based on mode of administration, the scores for the computer mode would be expected to be higher (or lower) than the paper-and pencil mode for both subject groups. If the scores are higher (or lower) on Administration 2 than Administration 1, the potential confound of an order effect would be examined. If Group 1 scored higher (or lower) than Group 2 on each Administration the composition of the subject groups would be examined. Normally, a control group would be used to examine these potential confounds. However, the counterbalanced design controls for these confounds by allowing the subjects assigned to experimental groups to act as their own control. Because no control group was required, the number of subjects available to be assigned to the experimental groups was increased. This factor was significant to this research due to the difficulty in retaining volunteers for a study which requires a retest.

The counterbalanced experimental design was used "to control for an apparent progressive change in the subject's response as he continues to serve in the experiment" (Townsend, 1953:64). As Townsend notes, practice and fatigue effects, known as "constant errors", must be controlled for, "or their effects will obscure the changes in the dependent variable which is being investigated" (1953:64-65). Additionally, "practice will probably produce varying amounts of improvement in the retest scores of different

individuals" (Anastasi, 1982:110-111). The retest effect typically is the result of subjects recalling their responses to the original test. Because of this memory confound, "the scores on the two administrations of the test are not independently obtained and the correlation between them will be spuriously high" (Anastasi, 1982:111).

Administration 1 and Administration 2 were controlled for time to reduce the retest effect. The mean time between administrations was 35 days with a standard deviation of 4.86 days. The range for all subjects was 22 to 49 days between administrations.

Subjects were randomly assigned to two experimental groups as shown in Table 6, "Counterbalanced Experimental Design". Group 1 (n = 50) completed Test 1 (computer) and Test 2 (paper-and-pencil) during Administration 1 (A_1) and completed Test 1 (paper-and-pencil) and Test 2 (computer) during Administration 2 (A_2). Group 2 (n = 40) completed Test 1 (paper-and-pencil) and Test 2 (computer) during A_1 and completed Test 1 (computer) and Test 2 (paper-and-pencil) during A_2.

McClave and Benson (1991) discuss the elements of a designed experiment. One element is the response, which is "the variable of interest in the experiment" (1991:860). The response is the dependent variable in the experiment. "Factors are those variables whose effect on the response is of interest to the experimenter" (1991:861). The factors are the independent variables for the experiment. McClave and Benson note that each factor can be comprised of multiple levels; these levels are the values which the factor can assume in the experiment. "Treatments are factor-level combinations utilized" in the experiment. (1991:861)

Independent Variables. The main factor of interest is the mode of test administration. The two levels of administration mode employed were paper-and-pencil and computer. Additionally, factors were employed to examine the effects of the time between administrations, the composition of the experimental groups, and the levels of the subjects' computer experience, on the response variables.

Dependent Variables. The response variables are the WPT raw score and the eleven VPI scale scores.

Computer System Requirements. The computer version of the WPT was the WPT-PC version 2.2 release. This software can be executed on IBM or IBM compatible hardware with 512K RAM, DOS 3.0 or higher, and one floppy disk drive. (Wonderlic Personnel Test - PC Version Manual, 1992:vi) The computer version of the VPI was the IBM version 3.0 release. This software can also be executed on IBM or IBM compatible hardware with 256K RAM, DOS 2.0 or greater, and two disk drives; one disk drive may be a hard disk. (Schrinka, 1989:4)

Both the WPT-PC and the VPI-Computer Version were completed and scored on IBM and IBM-compatible microcomputer systems. The WPT software was installed to and executed from a hard disk drive at AFIT and from a floppy disk drive at WSU. The VPI software was executed from floppy disk drives at both institutions with the results of the WPT being stored on the floppy diskette and the results of the VPI being stored on the hard disk drive.

TABLE 6

COUNTERBALANCED EXPERIMENTAL DESIGN

Test 1 - Wonderlic Personnel Test (Form IV)

Subject Group	Administration 1 (A_1)	Administration 2 (A_2)
Group 1 G_1	Computer Instrument	Paper-Pencil Instrument
Group 2 G_2	Paper-Pencil Instrument	Computer Instrument

Test 2 - Vocational Preference Inventory (VPI)

Subject Group	Administration 1 (A_1)	Administration 2 (A_2)
Group 1 G_1	Paper-Pencil Instrument	Computer Instrument
Group 2 G_2	Computer Instrument	Paper-Pencil Instrument

Statistical Analyses

After all data were collected and the instruments were scored, the following null hypotheses were tested:

Hypothesis 1 (Ho_1): There will be no significant difference in the WPT test scores:

Ho_{1A}: between the modes of administration (computer and paper),

Ho_{1B}: between the experimental groups (G_1 and G_2),

Ho_{1C}: between administrations (A_1 and A_2), nor

Ho_{1D}: between the subjects' reported computer experience levels.

Hypothesis 2 (Ho_2): There will be no significant difference in the VPI scale scores:

Ho_{2A}: between the modes of administration (computer and paper),

Ho_{2B}: between the experimental groups (G_1 and G_2),

Ho_{2C}: between administrations (A_1 and A_2), nor

Ho_{2D}: between the subjects' reported computer experience levels.

Statistical Procedures. Each instrument was scored and each individual WPT score and VPI scale score was recorded. Three-way analysis of variance was used to examine the experimental effects: mode of administration, order of administration, and experimental group composition; as well as the associated interaction of the experimental effects. The experimental effects were analyzed against each response (dependent) variable. Analyses of variance were also conducted to examine differences in response variables based on subjects' reported computer experience levels.

Analysis of variance was also performed on the response variables to examine the demographic effects. The demographic effects analyzed were, age, gender, educational institution, and educational level. These demographic items were self-reported on the pre-experimental questionnaire.

Subject reported preferences between the modes of administration were analyzed and reported.

IV. Analysis

This chapter presents the research results. SAS Institute version 6.06 software was executed on a VAX computer at the Air Force Institute of Technology to accomplish the statistical analyses. The chapter begins with a presentation of the results of the analyses of the experimental effects and the demographic effects on each response variable. Next, subject administration mode preferences, as reported in the post-experiment questionnaire, are presented. Finally, a summary of the results concludes the chapter.

Analysis of Experimental Effects

This section summarizes the differences in the response variables which can be attributed to the following experimental effects: mode of administration, order of administration, group composition, and subjects' reported computer experience levels.

WPT Experimental Effects Analysis. The response variable derived from the WPT is the raw test score. This score is based on the total number of correct responses from 50 possible items. Three-way analysis of variance was performed on this score to check for mode of administration effects, order effects, and group composition effects. The mode of administration and group composition differences were not significant.

The differences based on the order effect (Admin) were significant (Table 7). The overall mean score for the 90 subjects for Administration 1 (A_1) was 28.6 as compared to the overall mean score for Administration 2 (A_2) of 30.2 ($F = 5.94$; $p < .05$). Overall mean scores for the mode of administration were 29.8 for the paper-and-pencil version and 29.6 for the computer version. The overall mean score for Group 1 was 29.3 and the overall mean score for Group 2 was 30.2.

The interaction of the mode and group effects was also significant. This significance can be attributed to the experimental design. The interaction of the mode and group effects are an equivalent measure of the order of administration because the mode and group effects were counterbalanced across administrations. That is, the mode and group effects would have been confounded by the order administration had the effects not been counterbalanced.

TABLE 7

ANALYSIS OF VARIANCE
WPT SCORES

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Model	252.44	3	84.15	2.39	0.0704*
Error	6198.14	176	35.22		
Total	6450.58	179			
Admin	209.09	1	209.09	5.94	0.0158**
Mode	2.22	1	2.22	0.06	0.8020
Group	43.34	1	43.34	1.23	0.2688
Admin x Mode	41.13	1	41.13	1.17	0.2813
Admin x Group	0.01	1	0.01	0.00	0.9843
Mode x Group	206.88	1	206.88	5.87	0.0164**

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

Based on subjects' reported computer experience levels, a regression analysis was performed to examine the correlation between reported computer experience levels and the response variable (WPT score). The regression analysis yielded an *R-Square* of 0.0939. Based on the regression model it was determined that subjects' computer experience levels were not significantly predictive of their computer WPT scores.

VPI Experimental Effects Analyses. Each of the eleven VPI scale scores was examined using a three-way analysis of variance. The scales which had significant experimental effects are presented first, followed by a summary of the scales with nonsignificant experimental effects.

Investigative (I) Scale. This scale measures "a cluster of variables which include intellectuality, intelligence, unsociableness, scientism, and rationality" (Holland, 1985:7). Scores can range from 0 to 14 for this scale. Three-way analysis of variance was performed using the "I" scale score as the response variable (Table 8). The mode of administration and order of administration effects were shown to be nonsignificant. The overall scale mean score based on administration mode was 5.39 reported on paper-and-pencil administration and 5.87 reported on computer administration. The overall scale mean score based on administration order was 5.33 for A_1 and 5.92 for A_2.

The group effect was significant. The overall scale mean score for Group 1 was 6.38 and the overall scale mean score for Group 2 was 4.69 ($F = 6.27$; $p < .05$). The significance of this effect may be attributed to the composition of the groups. The interaction of the order effect and mode effect were also significant. The interaction of the order and mode effects is an equivalent measure of the group effect and may be attributed to the counterbalanced experimental design.

TABLE 8

ANALYSIS OF VARIANCE
VPI - INVESTIGATIVE SCALE

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Model	150.67	3	50.22	2.47	0.0631*
Error	3371.40	176	20.29		
Total	3722.06	179			
Admin	15.61	1	15.61	0.77	0.3817
Mode	10.27	1	10.27	0.51	0.4777
Group	127.31	1	127.31	6.27	0.0132**
Admin x Mode	124.79	1	124.79	6.13	0.0141**
Admin x Group	7.75	1	7.75	0.38	0.5375
Mode x Group	13.08	1	13.08	0.64	0.4231

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

Artistic (A) Scale. This scale measures subjects' "artistic interest, anxiety, expressiveness, originality, unconventionality, erratic effort and behavior" (Holland, 1985:7). Scores can range from 0 to 14 for this scale. Three-way analysis of variance was performed using the "A" scale score as the response variable (Table 9). The mode of administration and order of administration effects were not significant. The overall scale mean score based on administration mode was 3.30 reported on the paper-and-pencil

administration and 3.62 reported on the computer administration. The overall scale mean score based on administration order was 3.18 for A_1 and 3.73 for A_2.

The group effect was significant. The overall scale mean score for Group 1 was 4.03 and the overall scale mean score for Group 2 was 2.75 ($F = 6.14$; $p < .05$). The significance of this experimental effect can be attributed to the composition of the groups. The interaction of the order effect and mode effect were also significant at the .05 level. The interaction effect, as previously noted, can be attributed to the counterbalanced experimental design.

TABLE 9

ANALYSIS OF VARIANCE
VPI - ARTISTIC SCALE

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Model	89.28	3	29.76	2.51	0.0604*
Error	2087.45	176	11.86		
Total	2176.73	179			
Admin	13.34	1	13.34	1.12	0.2904
Mode	4.67	1	4.67	0.39	0.5311
Group	72.82	1	72.82	6.14	0.0142**
Admin x Mode	71.27	1	71.27	6.01	0.0152**
Admin x Group	3.12	1	3.12	0.26	0.6086
Mode x Group	11.79	1	11.79	0.99	0.3202

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

Infrequency (Inf) Scale. This scale measures "a cluster of positively correlated traits, attitudes aspirations, and deficiencies -- self deprecation, incompetency, socially undesirable traits, and a history of personal and vocational failure" (Holland, 1985:9). Scores can range from 0 to 20 for this scale. Three-way analysis of variance was performed using the "Inf" scale score as the response variable (Table 10). The mode of administration and order of administration effects were not significant. The overall scale mean score based on administration mode was 5.54 reported on paper-and-pencil administration and 5.65 reported on computer administration. The overall scale mean score based on administration order was 5.65 for A_1 and 5.54 for A_2.

The group effect was significant. The overall scale mean score for Group 1 was 5.32 and the overall scale mean score for Group 2 was 5.95 ($F = 6.14$; $p < .05$). The significance of this effect can be attributed to the composition of the groups. The interaction of the order effect and mode effect were also significant at the .10 level. The interaction effect can be attributed to the counterbalanced experimental design.

Acquiescence (Ac) Scale. This scale measures, at a low correlation level, a cluster of "sociability, dominance, dependence, impulsivity, cheerfulness, self-confidence, range of interest, conventionalism, and frankness" (Holland, 1985:9). Scores can range from 0 to 30 for this scale. Three-way analysis of variance was performed using the "Ac" scale score as the response variable (Table 11). The mode of administration and order of administration effects were not significant. The overall scale mean score based on administration mode was 11.02 reported on paper-and-pencil administration and 12.13 reported on computer administration. The overall scale mean score based on administration order was 11.51 for A_1 and 11.64 for A_2.

The group effect was significant. The overall scale mean score for Group 1 was 12.18 and the overall scale mean score for Group 2 was 10.83 ($F = 3.06$; $p < .10$). The significance of this effect can be attributed to the composition of the groups. The interaction of the order effect and mode effect were also significant at the .10 level. The interaction effect, as discussed earlier for the Wonderlic, can be attributed to the counterbalanced experimental design.

TABLE 10

ANALYSIS OF VARIANCE
VPI - INFREQUENCY SCALE

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Model	89.28	3	29.76	2.51	0.0604*
Error	2087.45	176	11.86		
Total	2176.73	179			
Admin	13.34	1	13.34	1.12	0.2904
Mode	4.67	1	4.67	0.39	0.5311
Group	72.82	1	72.82	6.14	0.0142**
Admin x Mode	71.27	1	71.27	6.01	0.0152**
Admin x Group	3.12	1	3.12	0.26	0.6086
Mode x Group	11.79	1	11.79	0.99	0.3202

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

TABLE 11

ANALYSIS OF VARIANCE
VPI - ACQUIESCENCE SCALE

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Model	137.16	3	45.72	1.71	0.1657
Error	4692.75	176	26.66		
Total	4829.91	179			
Admin	0.80	1	0.80	0.03	0.8627
Mode	55.56	1	55.56	2.08	0.1507
Group	81.60	1	81.60	3.06	0.0820*
Admin x Mode	80.81	1	80.81	3.03	0.0835*
Admin x Group	54.76	1	54.76	2.05	0.1536
Mode x Group	0.00	1	0.00	0.00	0.9897

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

Nonsignificant Findings. Analyses of variance on the remaining VPI scales revealed nonsignificant differences based on the effects analyzed. Appendix D presents the analysis of variance tables for each scale. The scales with nonsignificant differences were Realistic (R), Social (S), Enterprising (E), Conventional (C), Self-Control (Sc), Status (St), and Masculinity-Femininity (Mf). A complete summary of the means for each response variable by each experimental effect and the interactions is presented in Appendix E.

Based on subjects' reported computer experience levels, regression analyses were preformed to examine the correlation between reported computer experience levels and the VPI response variables (scale scores). The regression analyses, for the scales, yielded *R-Squares* that ranged from 0.0003 to 0.2387. Based on these regression models it was determined that subject's computer experience levels were not significantly predictive of their computer VPI scale scores.

Analysis of Demographic Effects

This section summarizes the differences in the response variables which are attributable to the following demographic effects: subject's gender, educational institution, education level, and age, respectively.

WPT Demographic Effects Analysis.

Gender. The differences in the WPT scores were not statistically significant at the .1 level. The WPT test manual reports that males score approximately 2 points higher than females for subjects with equivalent education levels. The mean score for males on the paper-and-pencil version was 30.3 and 28.1 for females ($F = 1.86$; $p > .10$). For the computer version, the mean for males was 30.2 and 27.7 for females ($F = 2.30$; $p > .10$). These means scores based on gender are consistent with the normative data presented in the test manual. The lack of significance in the present study may be attributed to the disproportionately small number of females in the sample which biases a statistical test in favor of significance.

Institution, Eduction, and Age. The differences in WPT scores based on the institution, education, and age effects were all significant at the .05 level for both modes of administration. All subjects at AFIT were graduate students and all subjects at

WSU were undergraduates. Because the WPT is a cognitive measure, the differences based on institution, education, and age were expected to be significant.

VPI Demographic Effects Analyses.

Gender. The following response variables differed significantly based on gender:

VPI Realistic Scale. This scale had significantly different mean scale scores based on gender at the .01 level. Both the paper-and-pencil ($F = 7.36$; $p < .01$) and the computer versions ($F = 8.38$; $p < .01$) were statistically significant. The mean scale score for males was 6.3 and for females was 3.2 on the computer version, and 5.4 for males and 2.7 for females on the paper-and-pencil version. This significant difference can be attributed to the scale's measure of interest and is consistent with the direction of gender differences reported in the VPI manual which shows that males consistently report higher levels.

VPI Conventional Scale. This scale had significantly different mean scale scores based on gender ($p < .10$). The computer version was statistically significant ($F = 3.67$; $p < .10$), while the paper-and-pencil version was not significant ($F = 1.08$; $p > .10$). The mean scale score for males was 4.3 and for females was 2.5 on the computer version, and 3.6 for males and 2.6 for females on the paper-and-pencil version. While there appears to be a difference based on the mode of administration, the differences attributable to mode effect were not significant. This difference may be attributable to the small sample size for females ($n = 21$). The mean scale scores are consistent with the direction of gender differences reported in the VPI manual (males report higher levels).

VPI Self-control Scale. This scale had significantly different mean scale scores based on gender ($p < .05$). Both the paper-and-pencil ($F = 6.13$; $p < .05$) and the computer ($F = 7.60$; $p < .01$) versions were statistically significant. The mean scale score for males was 7.9 and for females was 10.5 on the computer version, and 8.3 for males and 10.6 for females on the paper-and-pencil version. Because the VPI is a self-reported inventory, this significant difference can be attributed to the scale's measure of interest. The difference is consistent with the direction of gender differences reported in the VPI manual (females report higher levels).

VPI Masculinity-Femininity Scale. This scale had significantly different mean scale scores based on gender ($p < .01$). Both the paper-and-pencil and the computer versions were statistically significant. The mean scale score for males was 9.6 and for females was 7.1 on the computer version ($F = 18.54$; $p < .01$), and 9.8 for males and 6.7 for females on the paper-and-pencil version ($F = 21.71$; $p < .01$). This significant difference can be attributed to the scale's measure of interest. The normative data for this scale reveals that males typically report higher levels. This result is consistent with the direction of gender differences reported in the VPI manual.

VPI Status Scale. This scale had significantly different mean scale scores based on gender ($p < .01$). Both the paper-and-pencil ($F = 9.53$; $p < .01$) and the computer ($F = 8.90$; $p < .01$) versions were statistically significant. The mean scale score for males was 9.6 and for females was 7.7 on the computer version, and 9.6 for males and 7.6 for females on the paper-and-pencil version. This significant difference can be attributed to the scale's measure of interest, however, the VPI manual reports that females typically report higher levels than males. This difference may be attributed to the small

number of females in the subject population as noted previously. Additionally, this difference may be attributed to sampling error, that is, the sample of females in this study may be an atypical representation of females with regard to this scale.

VPI Infrequency Scale. This scale had significantly different mean scale scores based on gender ($p < .10$). Both the paper-and-pencil ($F = 6.82$; $p < .05$) and the computer ($F = 3.66$; $p < .10$) versions were statistically significant. The mean scale score for males was 5.4 and for females was 6.5 on the computer version, and 5.1 for males and 6.9 for females on the paper-and-pencil version. This difference is consistent with the direction of gender differences reported in the VPI manual (females report higher levels).

Nonsignificant findings. The remainder of the VPI scale differences based on gender were not significant at the .10 level for either mode of administration. The VPI Social scale was not significant on either mode of administration, however, the paper-and-pencil ($F = 2.79$; $p > .10$) mode of administration approached significance. The p-value for the computer administration ($F = 1.55$) was .2178. The mean scale scores for males was 3.8 and 4.8 for females on the computer version and 3.2 for males and 4.5 for females on the paper-and-pencil version. The higher levels reported by females are consistent with the VPI manual. The difference attributable to this scale based on the analysis of variance of the mode effect was not significant, therefore, the difference reported here may be attributed to either the sample size of females or sampling error.

TABLE 12

SUMMARY OF GENDER EFFECTS COMPARISONS

Dependent Variable	Male	Female	F	p
WPT - CV	30.2	27.7	2.30	0.1347
WPT - PV	30.3	28.1	1.86	0.1773
"R" - CV	6.3	3.2	8.38	0.0053***
"R" - PV	5.4	2.7	7.36	0.0086***
"I" - CV	6.0	5.3	0.50	0.4801
"I" - PV	5.5	5.9	0.45	0.5064
"A" - CV	3.5	3.9	0.27	0.6025
"A" - PV	3.2	3.7	0.81	0.3724
"S" - CV	3.8	4.8	1.55	0.2178
"S" - PV	3.2	4.5	2.79	0.1001
"E" - CV	5.1	3.8	2.19	0.1444
"E" - PV	4.7	3.4	1.73	0.1930
"C" - CV	4.3	2.5	3.67	0.0600*
"C" - PV	3.6	2.6	1.08	0.3035
"Sc" - CV	7.9	10.5	7.60	0.0077***
"Sc" - PV	8.3	10.6	6.13	0.0161**
"Mf" - CV	9.6	7.1	18.54	0.0001***
"Mf" - PV	9.8	6.7	21.71	0.0001***
"St" - CV	9.6	7.7	8.90	0.0041***
"St" - PV	9.6	7.6	9.53	0.0030***
"Inf" - CV	5.4	6.5	3.66	0.0604*
"Inf" - PV	5.1	6.9	6.82	0.0113**
"Ac" - CV	12.5	10.9	1.80	0.1841
"Ac" - PV	11.5	9.6	2.12	0.1507

Notes:
* = $p < .10$
** = $p < .05$
*** = $p < .01$

n = 90

PV = paper-and-pencil version
CV = computer version

Institution, Education, and Age. Tables 13-15 provide a summary of the response variable means and the associated statistical significance in differences based on the demographic effects, institution, education level, and age group, respectively.

Examination of the response variable means and the ANOVAs for each of these demographic effects revealed that the differences were consistent with the reported demographical data provided in the manuals. Some minor variances from the reported data in the manual were noted with regard to the VPI scales. These variances may be attributed to the subject population and the composition of the experimental groups. The majority of the subject population consisted of U.S. Air Force Officers (n = 75) who were enrolled in graduate programs. This portion of the population can be expected to represent a fairly homogeneous subject group with respect to reported occupational preferences. The remainder of the subject population (n = 15) was undergraduate students at WSU. While this subject group can be expected to be fairly heterogeneous with regard to reported occupational preferences, the small sample size of this group diminishes the effect of the heterogeneous group on the entire sample. Because the VPI is a self-reported preference inventory, the differences in the findings between this research and the normative demographical data, may be attributable to the relative homogeneity of the subject population. Additionally, the differences may be attributable to sampling error; the subject sample of this study may be an atypical representation of their respective populations as a whole.

The findings of this research were relatively consistent with the normative demographical data as reported in the WPT test manual. The homogeneity of the subject

population discussed above would not be expected to lead to a significant impact on the WPT response variable because the instrument is a cognitive test.

Additionally, it should be noted that an examination of the means and significance levels between the modes of administration were consistent with the analyses of variances for the associated experimental effect performed on each response variable.

TABLE 13

SUMMARY OF INSTITUTION EFFECTS COMPARISONS

Dependent Variable	INSTITUTION AFIT	WSU	F	p
WPT - CV	30.8	23.5	15.44	0.0002***
WPT - PV	31.2	22.9	21.02	0.0001***
"R" - CV	5.9	3.9	2.73	0.1036
"R" - PV	5.1	3.1	2.95	0.0912*
"I" - CV	5.9	5.9	0.00	0.9475
"I" - PV	5.4	5.3	0.00	0.9506
"A" - CV	3.6	3.7	0.00	0.9470
"A" - PV	3.3	3.1	0.09	0.7705
"S" - CV	3.8	5.4	3.20	0.0787*
"S" - PV	3.4	4.4	1.40	0.2408
"E" - CV	5.0	3.9	1.01	0.3200
"E" - PV	4.7	3.0	2.14	0.1490
"C" - CV	4.0	3.1	0.86	0.3584
"C" - PV	3.4	2.8	0.39	0.5323
"Sc" - CV	8.2	10.0	2.73	0.1034
"Sc" - PV	8.5	10.5	3.37	0.0711*
"Mf" - CV	9.4	7.1	12.31	0.0009***
"Mf" - PV	9.5	6.7	13.48	0.0005***
"St" - CV	9.5	7.3	10.08	0.0023***
"St" - PV	9.5	7.4	8.20	0.0057***
"Inf" - CV	5.4	6.7	3.71	0.0589*
"Inf" - PV	5.3	6.9	5.02	0.0288**
"Ac" - CV	12.3	11.5	0.35	0.5568
"Ac" - PV	11.3	9.8	0.99	0.3242

Notes:
* = $p < .10$
** = $p < .05$
*** = $p < .01$

n = 90

PV = paper-and-pencil version
CV = computer version

TABLE 14

SUMMARY OF EDUCATIONAL EFFECTS COMPARISONS

	EDUCATIONAL GROUPINGS						
Dependent Variable	**High Sch**	**High Sch +**	**Bach Deg.**	**Bach Deg +**	**Mast or +**	**F**	**p**
WPT - CV	20.4	25.0	31.1	30.5	31.5	4.33	0.0038***
WPT - PV	22.0	23.4	30.8	31.5	30.6	5.36	0.0009***
"R" - CV	4.6	3.6	5.5	5.8	6.7	0.88	0.4841
"R" - PV	4.0	2.7	3.5	5.4	6.3	1.78	0.1443
"I" - CV	5.4	6.2	4.5	5.5	8.8	2.02	0.1032
"I" - PV	3.2	6.4	3.2	4.5	8.0	3.51	0.0120**
"A" - CV	1.4	4.8	2.7	3.6	4.7	2.12	0.0888*
"A" - PV	0.4	4.5	2.4	3.5	3.9	3.40	0.0142**
"S" - CV	6.0	5.1	2.6	3.5	6.3	3.55	0.0114**
"S" - PV	4.4	4.4	2.9	3.0	5.2	1.66	0.1698
"E" - CV	6.6	2.6	4.5	5.0	5.5	1.37	0.2539
"E" - PV	4.8	2.1	4.2	4.8	4.8	0.96	0.4333
"C" - CV	6.0	1.6	3.6	3.7	5.5	2.08	0.0940*
"C" - PV	7.0	0.7	2.5	3.3	5.3	3.85	0.0074***
"Sc" - CV	10.8	9.6	8.4	8.3	7.7	0.84	0.5056
"Sc" - PV	10.8	10.3	8.7	8.4	8.6	0.89	0.4783
"Mf" - CV	7.0	7.2	9.2	9.6	9.1	3.26	0.0174**
"Mf" - PV	7.4	6.4	8.9	9.7	9.8	3.81	0.0078***
"St" - CV	6.2	7.8	9.3	9.4	10.3	3.25	0.0175**
"St" - PV	7.0	7.6	9.1	9.3	10.7	3.00	0.0252**
"Inf" - CV	8.2	6.0	5.8	5.6	4.5	2.23	0.0764*
"Inf" - PV	8.4	6.2	5.5	5.5	4.2	2.45	0.0557*
"Ac" - CV	11.8	11.3	11.2	11.9	14.8	1.32	0.2744
"Ac" - PV	9.4	10.0	10.2	10.8	14.2	1.51	0.2097

Notes:
* = $p < .10$
** = $p < .05$
*** = $p < .01$

n = 90

PV = paper-and-pencil version
CV = computer version

TABLE 15

SUMMARY OF AGE EFFECTS COMPARISONS

	AGE GROUPINGS							
Dependent Variable	**≤ 20**	**21-25**	**26-30**	**31-35**	**36-40**	**41-45**	**F**	**p**
WPT - CV	23.0	26.5	30.1	31.1	29.8	33.3	2.81	0.0240**
WPT - PV	22.6	25.5	31.0	30.8	30.5	32.3	3.52	0.0074***
"R" - CV	2.8	10.0	5.1	5.7	7.3	8.3	2.19	0.0672*
"R" - PV	2.5	11.5	4.2	4.6	6.2	8.5	2.80	0.0242**
"I" - CV	4.8	10.0	5.0	5.8	8.5	5.8	1.73	0.1412
"I" - PV	4.5	12.5	3.9	5.6	7.5	6.8	3.38	0.0092***
"A" - CV	2.7	3.0	4.2	2.9	4.0	6.3	1.69	0.1508
"A" - PV	2.1	4.0	3.5	3.2	2.8	7.0	2.66	0.0307**
"S" - CV	4.5	1.5	3.3	4.6	4.6	3.3	0.91	0.4824
"S" - PV	3.9	2.5	2.9	4.0	4.0	3.3	0.54	0.7485
"E" - CV	3.6	1.0	5.1	5.1	4.8	5.8	0.77	0.5777
"E" - PV	3.1	3.5	4.4	4.6	4.4	7.0	0.58	0.7144
"C" - CV	2.7	0.5	3.8	4.7	4.3	1.8	1.18	0.3487
"C" - PV	3.4	1.0	3.0	3.8	3.7	2.0	0.45	0.8148
"Sc" - CV	11.2	1.5	8.3	8.9	7.5	6.5	2.92	0.0200**
"Sc" - PV	11.1	1.5	8.9	8.6	8.6	7.8	2.36	0.0504*
"Mf" - CV	7.0	11.0	9.1	9.5	9.7	7.6	2.87	0.0217**
"Mf" - PV	6.5	11.6	9.3	9.3	9.7	9.5	2.64	0.0315**
"St" - CV	7.1	7.6	9.6	9.5	10.0	7.3	2.67	0.0300**
"St" - PV	7.4	8.0	9.4	9.2	10.2	8.6	1.66	0.1588
"Inf" - CV	7.1	4.6	5.8	5.2	5.2	6.3	1.27	0.2897
"Inf" - PV	7.3	4.6	5.9	4.9	4.8	6.3	1.75	0.1378
"Ac" - CV	10.0	14.0	11.9	12.3	14.2	11.5	1.05	0.3983
"Ac" - PV	9.3	16.5	10.2	10.8	13.1	14.3	1.56	0.1856

Notes:
* = $p < .10$
** = $p < .05$
*** = $p < .01$

n = 90

PV = paper-and-pencil version
CV = computer version

Post-Experiment Questionnaire

Question 1. Not counting this research study, how many psychological tests have you taken?

Six percent of the subjects reported the number of psychological tests they had taken to be "none". 18% of the subjects reported taking "one" previous test, 65% reported "more than one, but less than ten", and 10% reported taking "more than ten" tests.

Question 2. Not counting this research study, how many psychological tests have you taken in the computerized format?

74% of the subjects reported that the number of psychological tests they had taken in the computerized format to be "none". 11% of the subjects reported taking "one" computerized test and 14% reported taking "more than one, but less than ten" computerized tests. No subjects reported taking "more than ten" computerized tests.

Question 3. Not counting this research study, how many computerized questionnaires, surveys, or tests have you completed?

28% of the subjects reported completing "none". 16% of the subjects reported taking "one" previous computer instrument, 44% reported "more than one, but less than ten", and 12% reported taking "more than ten" tests.

Question 4. With regard to this research study, which test format did you prefer?

Seventeen percent of the subjects reported that they preferred the "Paper-and-Pencil" format with regard to this study. 54% of the subjects reported preferring the "computer" format, 26% reported that they "preferred Paper-and-Pencil on one test and Computer on the other test", and 3% reported "no preference" of test mode.

Question 5. Which version of the test(s) was/were easier to complete?

Thirteen percent of the subjects reported that the "Paper-and-Pencil" format was easier to complete with regard to this study. 50% of the subjects reported that the "computer" format was easier, 32% reported "Paper-and-Pencil easier on one test and Computer easier on the other test", and 5% reported "no difference" with regard to the test mode.

Question 6. If you could select which version to take on a future test, which would you choose?

Twenty-eight percent of the subjects reported that they would choose the "Paper-and-Pencil" format on a future test. 61% of the subjects reported they would choose the "computer" format and 11% reported that they had "no preference" as to test mode.

Summary of Results

This section summarizes the results of the research and interpretations of the significant findings by the experimental effects which were analyzed.

Between Mode of Administration Differences. The analyses of variance for all response variables revealed that the differences based on the mode of administration, paper-and-pencil compared with computer, were nonsignificant at the .10 level. Furthermore, none of the response variables approached significance with regard to this treatment effect. Table 16 summarizes the findings for each response variable based on the mode of administration.

TABLE 16

SUMMARY OF BETWEEN MODES OF ADMINISTRATION COMPARISONS

Dependent Variable	Paper Mean	Computer Mean	F	p
WPT	29.8	30.8	0.06	0.8020
VPI - Realistic	4.8	5.6	1.53	0.2175
VPI - Investigative	5.4	5.9	0.51	0.4777
VPI - Artistic	3.3	3.6	0.39	0.5311
VPI - Social	3.5	4.0	0.94	0.3336
VPI - Enterprising	4.4	4.8	0.50	0.4811
VPI - Conventional	3.3	3.9	0.85	0.3570
VPI - Self-Control	8.8	8.5	0.30	0.5874
VPI - Status	9.1	9.1	0.00	0.9751
VPI - Masculinity-Femininity	9.1	9.0	0.02	0.8910
VPI - Infrequency	5.5	5.7	0.10	0.7490
VPI - Acquiescence	11.0	12.1	2.08	0.1507

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

n = 90

Between Order of Administration Differences. The analyses of variance for response variables revealed a significant difference based on the order of administration, Administration 1 compared with Administration 2, between administrations of the Wonderlic Personnel Test. This difference was significant at the .05 level. The effect of the order of administration may be attributed to the fact that the WPT is a cognitive test. Due to the short time span between test administrations, the subjects' memory of the test items will result in improvement in the retest score (see e.g.; Anastasi, 1982:111). Differences based on order of administration for the remaining response variables were nonsignificant at the .10 level. No order of administration effect was expected with the VPI scales because the VPI is a self-reported measure. A memory effect or learning effect would not be expected to significantly effect the VPI response variables. Table 17 summarizes the findings for each response variable based on the order of administration.

TABLE 17

SUMMARY OF BETWEEN ORDER OF ADMINISTRATION COMPARISONS

Dependent Variable	Admin 1 Mean	Admin 2 Mean	F	p
WPT	28.6	30.8	5.94	0.0158**
VPI - Realistic	4.9	5.4	0.57	0.4504
VPI - Investigative	5.3	5.9	0.77	0.3817
VPI - Artistic	3.2	3.7	1.12	0.2904
VPI - Social	3.8	3.8	0.00	0.9664
VPI - Enterprising	4.6	4.6	0.00	0.9688
VPI - Conventional	3.6	3.6	0.01	0.9082
VPI - Self-Control	8.8	8.6	0.12	0.7272
VPI - Status	9.0	9.2	0.33	0.5527
VPI - Masculinity-Femininity	9.0	9.1	0.08	0.7840
VPI - Infrequency	5.7	5.5	0.10	0.7490
VPI - Acquiescence	11.5	11.6	0.03	0.8627

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

n = 90

Between Experimental Group Differences. The analyses of variance for response variables revealed a significant difference based on the experimental groups, Group 1 compared with Group 2, for the VPI Investigative, Artistic, Infrequency, and Acquiescence scales. These difference were significant at the .10 level. Differences based on the experimental groups for the remaining response variables were not significant. Table 18 summarizes the findings for each response variable based on the experimental group differences.

TABLE 18

SUMMARY OF BETWEEN EXPERIMENTAL GROUPS COMPARISONS

Dependent Variable	Group 1 Mean	Group 2 Mean	F	p
WPT	29.3	30.2	1.23	0.2688
VPI - Realistic	5.5	4.8	1.31	0.2536
VPI - Investigative	6.4	4.7	6.27	0.0132**
VPI - Artistic	4.0	2.8	6.14	0.0142**
VPI - Social	4.1	3.4	1.52	0.2186
VPI - Enterprising	4.3	5.0	1.23	0.2687
VPI - Conventional	3.8	3.4	0.43	0.5113
VPI - Self-Control	8.4	9.0	1.34	0.2480
VPI - Status	9.1	9.1	0.00	0.9944
VPI - Masculinity-Femininity	8.9	9.2	0.99	0.3199
VPI - Infrequency	5.3	6.0	3.26	0.0727*
VPI - Acquiescence	12.2	10.8	3.06	0.0820*

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

n = 90

Interaction of Experimental Effects. There were four statistically significant interaction effects at the .10 level based on the order and mode of administration interaction. Additionally, there was one statistically significant interaction effect at the .05 level based on the group and mode interaction. As previously noted, the significance

of these interaction effects can be attributed to the counterbalanced experimental design. Due to the counterbalance of the three experimental effects, a significant difference attributable to an individual effect will result in a significant difference attributable to the interaction of the remaining two effects. The interaction of the two remaining effects is an equivalent measure of the first effect. No other significant interaction effects were noted.

Between Reported Computer Experience Level Differences. The regression analyses for the response variables revealed that the subjects' reported computer experience levels were not significantly predictive of the subjects' computer WPT scores nor their computer VPI scale scores.

Subject Reported Mode of Administration Preferences. Based on the reported levels of subjects' preferences to the test mode of administration, 54% preferred the computerized mode of administration to the paper-and-pencil mode of administration. Additionally, 50% of the subjects reported that the computer mode of administration was "easier to complete" than the paper-and-pencil. It should be noted that 32% of the subjects felt that the computer mode was easier on one test and the paper-and-pencil mode was easier on the other. No data was collected to analyze which test was preferred in which mode of administration; however, it should be noted that a significant number of subjects had a mixed preference as to the mode of test administration. Only 13% of the subjects reported that the paper-and-pencil administration was "easier to complete".

V. Conclusions and Recommendations

This chapter summarizes the problem, the methodology, the findings, and the limitations of this research effort, presents conclusions, and provides recommendations for future research.

Summary

This research concerned the use of computer hardware and software as a means of human subject psychological data collection. An examination of computer instruments compared with paper-and-pencil instruments as methods of psychological data collection was the focus of this research effort. Specifically, differences in subject responses resulting from differences in the mode of instrument administration used in data collection was examined.

This research utilized a counterbalanced experimental design with subjects randomly assigned to two experimental groups. The counterbalanced design was utilized to control the potential confounding factors of group membership effects and order effects on the experimental factor. Subjects consisted entirely of volunteers from WSU and AFIT. Ninety subjects, consisting of 15 WSU students, and 75 AFIT students, completed all phases of the experiment. Of the 90 subjects, 21 were females and 69 were males. Eighty-three percent of the subjects had bachelor's degrees or higher and 17% were undergraduates.

Instruments utilized in the experiment were the Wonderlic Personnel Test (WPT) and the Vocational Preference Inventory (VPI). Computer and paper-and-pencil versions of each instrument were administered to subjects in the experiment.

The WPT is a twelve-minute timed cognitive ability test which contains 50 total items. The test contains quantitative, verbal, and spatial questions. Both versions of the WPT are scored based on the total number of correct responses made by the subject. This score was of specific relevance as a response variable in this research.

The VPI is a personality measurement instrument. The VPI is not timed and contains a listing of 160 occupations which the subject identifies to as being "appealing", "unappealing", or "indifferent". Both versions of the VPI are scored to yield eleven personality scale scores. These scale scores were of specific relevance as response variables in this research.

The reliability of each version of instrument utilized was computed. Specifically, internal consistency and test-retest reliabilities were of interest to this research. The spilt-half technique was utilized to examine the internal consistency of the instruments. Correlations of the two resulting halves were performed and the Spearman-Brown correction formula was applied. The resulting internal consistency coefficient for both computer and paper-and-pencil versions of the WPT and the paper-and-pencil version of the VPI were consistent with the normative data reported in the instrument manuals (Wonderlic, 1983; Holland, 1985). The internal consistency of the VPI computer version was not calculated due to an inability to collect individual item responses with the computer version.

Similarly, the test-retest coefficients of both instruments were consistent with the normative data reported in the manuals. Due to the counterbalanced experimental design, the test-retest reliabilities were controlled for mode of administration.

The primary experimental effect examined in this research was the mode of instrument administration, computer compared with paper-and-pencil. The experimental effects of order of administration, composition of experimental groups, and reported computer experience levels of subjects' were also examined. Demographic effects of gender, age, education levels, and subject's institution were also examined.

The pre- and post-experiment questionnaires were utilized to collect demographic data, data regarding each subject's computer experience level, and subjects' preference of administration mode.

Each instrument was administered to all subjects in both versions. Administration 1 and Administration 2 were controlled for time to reduce the retest effect. Mean time between administrations was 35 days with a standard deviation of 4.86 days. The range of time between administrations for all subjects was 22 to 49 days.

Data were collected to examine the following null hypotheses:

Ho_1: There will be no significant difference in the WPT test scores:

Ho_{1A}: between modes of administration,

Ho_{1B}: between experimental groups,

Ho_{1C}: between administrations, nor

Ho_{1D}: between the computer experience levels reported by the groups.

Ho_2: There will be no significant difference in the VPI scale scores:

Ho_{2A}: between modes of administration,

Ho_{2B}: between experimental groups,

Ho_{2C}: between administrations, nor

Ho_{2D}: between the computer experience levels reported by the groups.

Three-way analysis of variance (ANOVA) was utilized to analyze the main experimental effect of mode of administration. Additionally, ANOVAs were used to examine the experimental effects, order of administration, group composition, and subjects' reported computer experience levels. Finally, ANOVAs were used to examine the differences in the response variables attributable to the demographic effects of gender, age, education, and institution.

Limitations of the Study

Certain limitations apply to this research effort. First, although the assignment of subjects to the experimental groups was random in nature, the groups consisted of predominantly more males ($n = 69$) than females ($n = 21$). The experimental groups also consisted primarily of well-educated subjects. In fact, 83% of the subject population possessed a bachelor's degree or higher.

Another potential experimental group bias is the fact that the majority of the group was male military officers. Group composition may have had some biasing effect on the sample and the experiment. Also, the original subject population consisted of 40 volunteer students from WSU and 87 from AFIT. Of the 40 original WSU volunteers, only 15 completed all phases of the experiment. Similarly, only 75 of the original 87 AFIT volunteers completed all phases of the experiment. The small sample size of female participants, coupled with the smaller than expected number of subjects from WSU limit the power of the statistical tests. Consideration of these limitations is recommended when drawing conclusions from the experiment.

Conclusions

Most of the results of this research with regard to the experimental effects examined were not significant. With regard to the main experimental effect, mode of administration, all of the null hypotheses (Ho_{1A} and Ho_{2A}) were not rejected. There was no significant difference in the response variables between computer and paper-and-pencil modes of administration.

With regard to the experimental effect, experimental group composition, all null hypotheses (Ho_{1B} and Ho_{2B}), except four VPI scales, were not rejected. There were significant differences between the experimental groups on the VPI Investigative, Artistic, Infrequency, and Acquiescence scale scores. All other response variables were not significantly different relative to experimental group composition.

With regard to the experimental effect, order of administration, all null hypotheses (Ho_{2C}) related to the eleven VPI scales were not rejected. That is, there was no significant difference between Administration 1 and Administration 2 in subject's reported levels related to the eleven VPI scales. The null hypothesis (Ho_{1C}) related to the WPT response variable was rejected. That is, there was a significant difference between WPT scores between the administrations. This factor was controlled for mode of administration due to the experimental design. The subjects scored significantly higher on Administration 2 over Administration 1 of this response variable, regardless of mode of administration.

With regard to the experimental effect, subject's reported computer experience level, the null hypotheses (Ho_{1D} and Ho_{2D}) were not rejected. The regression analyses revealed that this factor was not significantly predictive of the response variables.

Data collected utilizing the post-experiment questionnaire revealed 54% of participating subjects preferred the computer mode over the paper-and-pencil mode of administration of the experimental instruments. Also, 50% of the subjects reported the computer version to be "easier to complete" than the paper-and-pencil version of the instruments. 32% of the subjects reported the computer version of one instrument easier to complete and the paper-and-pencil version of the other instrument easier to complete. Only 13% of subjects reported to prefer the paper-and-pencil version as "easier to complete" than the computer version of the instruments.

Based on the results of this research, the following conclusions can be drawn:

1. The microcomputer can provide a reliable means of administering, recording, and scoring psychological tests. There were no significant differences in the response variables between the modes of administration.

2. Taking psychological tests on the computer is preferable to paper-and-pencil administration. Subjects reported that they preferred the computer version over the paper-and-pencil version and that the computer version was easier to complete. Furthermore, subjects reported that, if given the option, in the future they would choose to take computer-based tests over paper-and-pencil versions.

3. The use of standardized norms in the design and application of the two psychological measures used in this research was effective. The computer and paper-and-pencil versions of each measure were designed and scored alike; therefore, the norms associated with the scoring of each measure were alike. The fact that no significant differences were found between the modes of administration of either of the measures

lends validity to the intended use of the measures utilized and to the use of standardized norms in computer and paper-and-pencil versions of the cited measures.

Recommendations for Future Research

As noted in the previous section, 50% of the subjects reported that the computer version was "easier to complete" than the paper-and-pencil administrations of the instruments. However, 32% of the subjects reported that the paper-and-pencil version was easier to complete on one instrument while the computer version was easier to complete on the other instrument. No data were collected to determine which instrument the subjects felt was easier by which mode. It is apparent that a substantial percentage (32%) of the subjects felt that one instrument was more suitable to computer administration. Therefore, it is recommended that future research be conducted to examine which types of instruments are more suitable to computer applications. This research did not examine the correlation of responses to subjects' preferred mode of administration. Further study of preferred format of types of measures and resulting responses is therefore recommended.

The literature reviewed also suggest that careful consideration should be given to the types of instruments which are administered in the computer mode. (Elithorn *et al.*, 1982:247,249; Space 1981:603; Thompson and Wilson, 1982:287; Volans, 1982:301-304). Although this research supported the use of standardized norms for both the computer and paper-and-pencil versions of the instruments utilized, more research in the use of standardized norms among measures is recommended, specifically, examination of the use of standardized norms relative to the instrument's administration complexity. A proposed hypothesis is: Are there significantly increased differences among the

standardized paper-and-pencil norms and the computer norms as the level of test administration complexities increase? As tests are developed specifically for computer administration, the use of existing standardized norms will diminish. New normative data will have to be collected as the new instruments are developed.

Another potentially powerful application of the computer, noted in the literature, is the use of adaptive testing. The speed and storage capabilities of the computer will allow complex adaptive testing instruments to be developed. Further research into the normative data and instrument reliabilities of these applications is essential to acceptance and implementation of the instruments.

Based on the limitations of this research, it would be beneficial to further examine the effect of mode of administration on response variables employing a more diverse subject population. The subject population for this research consisted of, primarily, well-educated and computer-literate subjects. Further research into subject populations which are more balanced with regard to computer literacy would be worthwhile. Additionally, future research in this area should attempt to reduce the gender bias experienced in this study.

For future studies in this area, the counterbalanced experimental design is recommended due to the ability to control for the confounding factors of order effect and experimental group composition effect. This allowed these factors to be controlled without the assignment of subjects to control groups. Due to the difficulty in obtaining and retaining subject for a study of this nature, the number of subjects which can be assigned to experimental groups is a critical experimental design factor. This study

demonstrated the effectiveness of utilizing the counterbalanced experimental design to control for these confounding factors.

This research focused on a comparison of paper-and-pencil and computer administration modes in psychological testing. Specifically, differences in the mode of administration of the WPT and the VPI were examined. While this research revealed no significant differences between the modes of administration with regard to these instruments, continued examination into applications of computer technologies is necessary. This continued research is not only important in the field of psychology, but also in any field where computer technologies are being introduced. With the reduction in the cost of computers, coupled with the subsequent increases in computer power and availability, introduction of the computer into almost every area of our lives is inevitable. Continued research of this subject is essential to ensuring that the application of the new computer technologies is both effective and beneficial.

Appendix A: Pre-experiment Questionnaire

FOR EACH QUESTION, CIRCLE THE MOST APPROPRIATE RESPONSE.

001. Indicate your gender.

1. Male.
2. Female.

002. Find the number below that includes your age.

1. 20 or less.
2. 21 to 25.
3. 26 to 30.
4. 31 to 35.
5. 36 to 40.
6. 41 to 45.
7. 46 or more.

003. Indicate the highest level of education you have achieved.

1. High school diploma or graduate equivalent diploma.
2. Some courses beyond high school but no degree awarded.
3. Associate's Degree.
4. Bachelor's Degree.
5. Bachelor's Degree, with some graduate courses.
6. Master's Degree or higher.

004. Do you have access to a computer on a regular basis?

1. No.
2. I don't know.
3. Yes.

005. Do you have a computer at home?

1. No.
2. Yes.

006. How often do you use a computer at work or school?

1. Never.
2. Occasionally.
3. About half of the time.
4. Frequently.
5. Most of the time.

007. How long have you used a computer at work, school, or home?

1. Never used computer.
2. Some experience, but less than 1 1/2 years.
3. 1 1/2 to 3 years.
4. 3 to 5 years.
5. 5 to 10 years.
6. 10 years or more.

008. How much confidence do you have using your computer to meet the needs of your job or school work?

1. Very low confidence.
2. Low confidence.
3. Moderate confidence.
4. High confidence.
5. Very high confidence.

009. Indicate the highest level of computer training you have completed.

1. None.
2. High school course(s).
3. Adult, continuing education, or college course(s).
4. Associate Degree in Computer Science, Information Resource Management or Management Information Systems.
5. Bachelor Degree in Computer Science, Information Resource Management or Management Information Systems.
6. Master Degree in Computer Science, Information Resource Management or Management Information Systems.
7. Other (fill in on this form) _____

010. What percentage of your computer knowledge was self taught.

1. Less than 25%.
2. 26 to 50%.
3. 51 to 75%.
4. More than 75%.

For Questions 011 through 018, rate how confident you feel using that application. Use the following scale for each question.

1. Don't use it at all.
2. Don't recognize this.
3. I can perform only basic functions following prompts or menus; I usually need help recovering from mistakes.
4. I can perform all of the basic functions and follow instructions in the manual for more advanced functions; I sometimes require help in performing the more advanced functions.
5. I can perform all of the basic and advance functions; I rarely, if ever, require assistance.
6. I can perform all of the functions of the application; others seek my help in using the application.

011. Databases.

012. Word Processing.

013. Spreadsheets.

014. Graphics.

015. Telecommunications.

016. CAD or CAM.

017. DOS or equivalent commands.

018. Windows applications.

(Adapted from Casey and Kveene, 1991:87-90).

Appendix B: Post-experiment Questionnaire

FOR EACH QUESTION, CIRCLE THE MOST APPROPRIATE RESPONSE.

001. Not counting this research study, how many psychological tests have you taken.

1. None.
2. One.
3. More than one, but less than ten.
4. More than ten.

002. Not counting this research study, how many psychological tests have you taken in the computerized format.

1. None.
2. One.
3. More than one, but less than ten.
4. More than ten.

003. Not counting this research study, how many computerized questionnaires, surveys, or tests have you completed.

1. None.
2. One.
3. More than one, but less than ten.
4. More than ten.

004. With regard to this research study, which test format did you prefer.

1. Paper-and-Pencil.
2. Computer.
3. Preferred Paper-and-Pencil on one test and Computer on the other test.
4. No Preference.

005. Which version of the test(s) was/were easier to complete.

1. Paper-and-Pencil.
2. Computer.
3. Paper-and-Pencil easier on one test and Computer easier on the other test.
4. No difference.

006. If you could select which version to take on a future test, which would you choose?

1. Paper-and-Pencil.
2. Computer.
3. No preference.

Appendix C: Clinical Interpretation of VPI Scales

1. Realistic (R) Scale. High scorers regard themselves as practical-minded and normal people. Their hardheaded orientation is consistent with their mechanical skills and interests and their lack of skill in interpersonal relations. They show low social interests and an aversion for problems requiring sensitivity to one's own feelings, or those of others, as in the arts or persuasive roles.

2. Investigative (Intellectual) (I) Scale. High scorers are concerned with science, mathematics, and theory. They prefer to "think through" problems rather than "act out" problems. They value science and aesthetic problems and deprecate social, political, and business activities. High scorers tend to be bright, scholarly, persistent, and to have high educational aspirations.

3. Artistic (A) Scale. High scorers have artistic, musical, and literary interests. They resemble the stereo-type of the artist in some ways -- they may be immature, anxious, sensitive, and feminine. They tend to be original, imaginative, complex, unconventional, and introverted.

4. Social (S) Scale. High scorers have social interests and prefer teaching or therapeutic roles. They are responsible, accepting of feminine impulses and roles, and facile and insightful in interpersonal relationships. High scorers have good role playing ability and the ability to relate to others, or the ability to form "close" as opposed to "superficial" relationships.

5. Enterprising (E) Scale. High scorers tend to be dominant, sociable, cheerful, and adventurous. They differ from Conventional high scorers in their need for ambiguous verbal tasks rather than structured activity, and a greater need for power. They are similar to Social high scorers, but more apt to be persuasive than helpful. This scale is, in one sense, an activity scale which represents euphoric behavior at one extreme and depressive behavior at the other. High scorers prefer social interaction as medium of personal expression, but dislike well-defined language or work situations. They conceive of themselves as strong leaders, regard their verbal and persuasive skills as their greatest assets, and have strong need to achieve and secure high status.

6. Conventional (C) Scale. High scorers are conventional, conforming, status-oriented, ethnocentric, unoriginal. They have introcepted the culture with unusual completeness and often appear controlled and defensive. They prefer structured rote verbal and numerical

activities and generally prefer subordinate roles. They seem to achieve their goals by conforming, living by the rules, and ordering their lives. In this fashion, they obtain satisfaction and simultaneously avoid conflict and anxiety which appear to be aroused by ambiguous situations and problems of interpersonal relationships. Their habitual subordination of their personal needs appears to make them generally productive and effective in will-structured tasks. Their values and attitudes include strong identifications with power, money, and status.

7. Self-Control (Sc) Scale. Self-control is defined simply as the habitual inhibition of impulses to act out in behavior, thinking, or fantasy. In the words of the man on the street, it is captured in the expressions, "so and so is careful, smooth," "always says the right thing," "never makes anyone mad," "stays out of trouble."

High scores indicate overcontrol. High scorers are often described as inhibited, constricted, passive, and responsible. High scores indicate concern with physical injury, illness, preoccupation with physical and medical problems, and potentially dangerous or threatening physical situations.

Low scores indicated impulsiveness and a tendency to "act out," which is suggestive of asocial psychopathy. Average scores are associated with a healthy spontaneity in living and originality, when associated with other positive signs.

8. Masculinity-Femininity (Mf) Scale. High scores indicate frequent choice of traditionally masculine occupation roles; that is, choices commonly preferred by men. Low scores indicate occupations traditionally preferred by women. This scale can be used to estimate the degree to which these traditional sex-typing has been incorporated into a person's thinking about occupations. For example, a man with a low score is a more likely prospect for occupations dominated by women than a man with a high score. Likewise, a woman with a high score is more likely prospect for occupations dominated by men. The Mf scale is also useful for detecting faking, because of its correlations with interest scales. For example, high S and A scores should go with a low Mf score.

9. Status (St) Scale. High scores are indicative of vocational choices with high prestige ranking. Generally, individual scores are positively correlated with the person's social origin. Scores appear to represent a measure of the person's expectation of, and need for, status or prestige. They may also represent a crude measure of the need for upward mobility. The St scale provides an estimate of self-esteem and self-confidence; that is, self-confidence is associated with high scores, self-deprecation wit low scores.

10. Infrequency (Inf) Scale. The scored items include preferences for unpopular, female-dominated, low status occupations and the rejection of male-dominated high status, popular occupations requiring various kinds of interpersonal, artistic, and intellectual talent. This analysis suggests that high scorers have atypical vocational preferences and, by implication, self-deprecating attitudes and deviant attitudes about their culture.

In contrast, low scorers see the occupational world in a popular way, have positive evaluations of their abilities and personalities, and high aspirations.

11. Acquiescence (Ac) Scale. People who prefer many occupations are expressing sociable, cheerful, active, frank, and conventional outlooks about the vocational, world, whereas people who like only a few occupations are expressing an unsociable, depressive, and unconventional outlook. Having many preferences is also associated with self-confidence; having few preferences is associated with self-deprecation. Extremely high Ac scores are associated with poor judgement, lack of personal integration. In addition, high scorers tend to be of two types. One type is characterized by poor judgement, lack of personal integration, and hyperactivity. Another type is well-integrated and has multiple interests and talents. A review of non-VPI information (work history, current aspirations, current level of work or academic achievement, vocational identity, and other personal data) usually clarifies the probable meaning of an elevated profile. (Holland, 1985:6-9)

Appendix D: Nonsignificant ANOVA Tables

TABLE 19

ANALYSIS OF VARIANCE
VPI - REALISTIC SCALE

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Model	60.76	3	20.25	1.08	0.3601
Error	3308.82	176	18.80		
Total	3369.58	179			
Admin	10.76	1	10.76	0.57	0.4504
Mode	28.80	1	28.80	1.53	0.2175
Group	24.67	1	24.67	1.31	0.2536
Admin x Mode	21.20	1	21.20	1.13	0.2897
Admin x Group	25.33	1	25.33	1.35	0.2473
Mode x Group	7.29	1	7.29	0.39	0.5343

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

TABLE 20

ANALYSIS OF VARIANCE
VPI - SOCIAL SCALE

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Model	30.88	3	10.29	0.82	0.4828
Error	2201.10	176	12.51		
Total	2231.98	179			
Admin	0.02	1	0.02	0.00	0.9644
Mode	11.76	1	11.76	0.94	0.3336
Group	19.07	1	19.07	1.52	0.2186
Admin x Mode	19.10	1	19.10	1.53	0.2182
Admin x Group	11.79	1	11.79	0.94	0.3330
Mode x Group	0.05	1	0.05	0.00	0.9475

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

TABLE 21

ANALYSIS OF VARIANCE
VPI - ENTERPRISING SCALE

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Model	25.18	3	8.39	0.58	0.6280
Error	2541.60	176	14.44		
Total	2566.78	179			
Admin	0.02	1	0.02	0.00	0.9688
Mode	7.20	1	7.20	0.50	0.4811
Group	17.78	1	17.78	1.23	0.2687
Admin x Mode	17.96	1	17.96	1.24	0.2663
Admin x Group	7.38	1	7.38	0.51	0.4756
Mode x Group	0.20	1	0.20	0.01	0.9059

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

TABLE 22

ANALYSIS OF VARIANCE
VPI - CONVENTIONAL SCALE

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Model	19.31	3	6.44	0.43	0.7327
Error	2641.90	176	15.01		
Total	2661.20	179			
Admin	0.20	1	0.20	0.01	0.9082
Mode	12.80	1	12.80	0.85	0.3570
Group	6.50	1	6.50	0.43	0.5113
Admin x Mode	6.31	1	6.31	0.42	0.5178
Admin x Group	12.60	1	12.60	0.84	0.3608
Mode x Group	0.00	1	0.00	0.00	0.9897

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

TABLE 23

ANALYSIS OF VARIANCE
VPI - SELF-CONTROL SCALE

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Model	25.41	3	8.47	0.57	0.6325
Error	2594.60	176	14.74		
Total	2620.00	179			
Admin	1.80	1	1.80	0.12	0.7272
Mode	4.36	1	4.36	0.30	0.5874
Group	19.80	1	19.80	1.34	0.2480
Admin x Mode	19.25	1	19.25	1.31	0.2547
Admin x Group	3.80	1	3.80	0.26	0.6122
Mode x Group	1.25	1	1.25	0.08	0.7715

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

TABLE 24

ANALYSIS OF VARIANCE
VPI - STATUS SCALE

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Model	2.01	3	0.67	0.12	0.9492
Error	997.52	176	5.67		
Total	999.53	179			
Admin	2.01	1	2.01	0.35	0.5527
Mode	0.01	1	0.01	0.00	0.9751
Group	0.00	1	0.00	0.00	0.9944
Admin x Mode	0.00	1	0.00	0.00	0.9863
Admin x Group	0.01	1	0.01	0.00	0.9721
Mode x Group	2.01	1	2.01	0.35	0.5526

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

TABLE 25

ANALYSIS OF VARIANCE
VPI - MASCULINITY-FEMININITY SCALE

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Model	5.18	3	1.73	0.37	0.7775
Error	830.46	176	4.72		
Total	835.64	179			
Admin	0.36	1	0.36	0.08	0.7840
Mode	0.09	1	0.09	0.02	0.8910
Group	4.69	1	4.69	0.99	0.3199
Admin x Mode	4.74	1	4.74	1.00	0.3176
Admin x Group	0.13	1	0.13	0.03	0.8661
Mode x Group	0.40	1	0.40	0.09	0.7710

Notes: * = $p < .10$
** = $p < .05$
*** = $p < .01$

Appendix E: Summary of Means

This appendix contains a summary of the means for each response variable as obtained from the analysis of variance procedures. The tables provide the mean values for each response variable based on order of administration, mode of administration, and group composition. Included are the two-way and three-way interactions of each effect. For each effect, there are two levels which are represented numerically by a "1" or "2".

For the experimental effect "ADMIN" (order of administration), the "1" represents Administration 1 and the "2" represents Administration 2. For the experimental effect "MODE" (administration mode), the "1" represents paper-and-pencil administration and the "2" represents computer administration. For the experimental effect "GROUP" (group composition), the "1" represents Group 1 and the "2" represents Group 2.

TABLE 26

SUMMARY OF MEANS
WPT SCORES

Level of ADMIN	N	WO Mean	WO SD
1	90	28.6111111	5.62437401
2	90	30.7666667	6.20447464

Level of MODE	N	WO Mean	WO SD
1	90	29.8000000	6.08590193
2	90	29.5777778	5.95106973

Level of GROUP	N	WO Mean	WO SD
1	100	29.2500000	6.57263227
2	80	30.2375000	5.19309120

Level of ADMIN	Level of MODE	N	WO Mean	WO SD
1	1	40	29.1500000	4.93314275
1	2	50	28.1800000	6.13684091
2	1	50	30.3200000	6.87628281
2	2	40	31.3250000	5.27882806

Level of ADMIN	Level of GROUP	N	WO Mean	WO SD
1	1	50	28.1800000	6.13684091
1	2	40	29.1500000	4.93314275
2	1	50	30.3200000	6.87628281
2	2	40	31.3250000	5.27882806

Level of MODE	Level of GROUP	N	WO Mean	WO SD
1	1	50	30.3200000	6.87628281
1	2	40	29.1500000	4.93314275
2	1	50	28.1800000	6.13684091
2	2	40	31.3250000	5.27882806

Level of ADMIN	Level of MODE	Level of GROUP	N	WO Mean	WO SD
1	1	2	40	29.1500000	4.93314275
1	2	1	50	28.1800000	6.13684091
2	1	1	50	30.3200000	6.87628281
2	2	2	40	31.3250000	5.27882806

TABLE 27

SUMMARY OF MEANS
VPI - REALISTIC SCALE SCORES

Level of ADMIN	N	VR Mean	VR SD
1	90	4.94444444	4.12272712
2	90	5.43333333	4.55441507

Level of MODE	N	VR Mean	VR SD
1	90	4.78888889	4.20404953
2	90	5.58888889	4.45676974

Level of GROUP	N	VR Mean	VR SD
1	100	5.52000000	4.52709018
2	80	4.77500000	4.08137170

Level of ADMIN	Level of MODE	N	VR Mean	VR SD
1	1	50	4.94000000	4.24941773
1	2	40	4.95000000	4.01248053
2	1	40	4.60000000	4.19278990
2	2	50	6.10000000	4.76059506

Level of ADMIN	Level of GROUP	N	VR Mean	VR SD
1	1	50	4.94000000	4.24941773
1	2	40	4.95000000	4.01248053
2	1	50	6.10000000	4.76059506
2	2	40	4.60000000	4.19278990

Level of MODE	Level of GROUP	N	VR Mean	VR SD
1	1	50	4.94000000	4.24941773
1	2	40	4.60000000	4.19278990
2	1	50	6.10000000	4.76059506
2	2	40	4.95000000	4.01248053

Level of ADMIN	Level of MODE	Level of GROUP	N	VR Mean	VR SD
1	1	1	50	4.94000000	4.24941773
1	2	2	40	4.95000000	4.01248053
2	1	2	40	4.60000000	4.19278990
2	2	1	50	6.10000000	4.76059506

TABLE 28

SUMMARY OF MEANS
VPI - INVESTIGATIVE SCALE SCORES

Level of ADMIN	N	VI Mean	VI SD
1	90	5.33333333	4.31875971
2	90	5.92222222	4.79519370

Level of MODE	N	VI Mean	VI SD
1	90	5.38888889	4.44616852
2	90	5.86666667	4.68370352

Level of GROUP	N	VI Mean	VI SD
1	100	6.38000000	4.99045554
2	80	4.68750000	3.78067622

Level of ADMIN	Level of MODE	N	VI Mean	VI SD
1	1	50	5.90000000	4.76488002
1	2	40	4.62500000	3.62107436
2	1	40	4.75000000	3.97911213
2	2	50	6.86000000	5.20991989

Level of ADMIN	Level of GROUP	N	VI Mean	VI SD
1	1	50	5.90000000	4.76488002
1	2	40	4.62500000	3.62107436
2	1	50	6.86000000	5.20991989
2	2	40	4.75000000	3.97911213

Level of MODE	Level of GROUP	N	VI Mean	VI SD
1	1	50	5.90000000	4.76488002
1	2	40	4.75000000	3.97911213
2	1	50	6.86000000	5.20991989
2	2	40	4.62500000	3.62107436

Level of ADMIN	Level of MODE	Level of GROUP	N	VI Mean	VI SD
1	1	1	50	5.90000000	4.76488002
1	2	2	40	4.62500000	3.62107436
2	1	2	40	4.75000000	3.97911213
2	2	1	50	6.86000000	5.20991989

TABLE 29

SUMMARY OF MEANS
VPI - ARTISTIC SCALE SCORES

Level of ADMIN	N	VA Mean	VA SD
1	90	3.18888889	3.26676[illegible]
2	90	3.73333333	3.6926894[illegible]

Level of MODE	N	VA Mean	VA SD
1	90	3.30000000	3.28240115
2	90	3.62222222	3.69201318

Level of GROUP	N	VA Mean	VA SD
1	100	4.03000000	3.80444981
2	80	2.75000000	2.91439031

Level of ADMIN	Level of MODE	N	VA Mean	VA SD
1	1	50	3.64000000	3.50370067
1	2	40	2.62500000	2.88841630
2	1	40	2.87500000	2.97155316
2	2	50	4.42000000	4.08126631

Level of ADMIN	Level of GROUP	N	VA Mean	VA SD
1	1	50	3.64000000	3.50370067
1	2	40	2.62500000	2.88841630
2	1	50	4.42000000	4.08126631
2	2	40	2.87500000	2.97155316

Level of MODE	Level of GROUP	N	VA Mean	VA SD
1	1	50	3.64000000	3.50370067
1	2	40	2.87500000	2.97155316
2	1	50	4.42000000	4.08126631
2	2	40	2.62500000	2.88841630

Level of ADMIN	Level of MODE	Level of GROUP	N	VA Mean	VA SD
1	1	1	50	3.64000000	3.50370067
1	2	2	40	2.62500000	2.88841630
2	1	2	40	2.87500000	2.97155316
2	2	1	50	4.42000000	4.08126631

TABLE 30

SUMMARY OF MEANS
VPI - SOCIAL SCALE SCORES

Level of ADMIN	N	VS Mean	VS SD
1	90	3.77777778	3.49888514
2	90	3.80000000	3.58273011

Level of MODE	N	VS Mean	VS SD
1	90	3.53333333	3.36939297
2	90	4.04444444	3.68693750

Level of GROUP	N	VS Mean	VS SD
1	100	4.08000000	3.65059840
2	80	3.42500000	3.36314726

Level of ADMIN	Level of MODE	N	VS Mean	VS SD
1	1	50	3.84000000	3.50136999
1	2	40	3.70000000	3.53879597
2	1	40	3.15000000	3.19895816
2	2	50	4.32000000	3.81425896

Level of ADMIN	Level of GROUP	N	VS Mean	VS SD
1	1	50	3.84000000	3.50136999
1	2	40	3.70000000	3.53879597
2	1	50	4.32000000	3.81425896
2	2	40	3.15000000	3.19895816

Level of MODE	Level of GROUP	N	VS Mean	VS SD
1	1	50	3.84000000	3.50136999
1	2	40	3.15000000	3.19895816
2	1	50	4.32000000	3.81425896
2	2	40	3.70000000	3.53879597

Level of ADMIN	Level of MODE	Level of GROUP	N	VS Mean	VS SD
1	1	1	50	3.84000000	3.50136999
1	2	2	40	3.70000000	3.53879597
2	1	2	40	3.15000000	3.19895816
2	2	1	50	4.32000000	3.81425896

TABLE 31

SUMMARY OF MEANS
VPI - ENTERPRISING SCALE SCORES

Level of ADMIN	N	VE Mean	VE SD
1	90	4.62222222	3.67676512
2	90	4.60000000	3.91424939

Level of MODE	N	VE Mean	VE SD
1	90	4.41111111	3.65013296
2	90	4.81111111	3.92884591

Level of GROUP	N	VE Mean	VE SD
1	100	4.33000000	3.71199258
2	80	4.96250000	3.87279950

Level of ADMIN	Level of MODE	N	VE Mean	VE SD
1	1	50	4.16000000	3.46622184
1	2	40	5.20000000	3.89081759
2	1	40	4.72500000	3.88941694
2	2	50	4.50000000	3.97055489

Level of ADMIN	Level of GROUP	N	VE Mean	VE SD
1	1	50	4.16000000	3.46622184
1	2	40	5.20000000	3.89081759
2	1	50	4.50000000	3.97055489
2	2	40	4.72500000	3.88941694

Level of MODE	Level of GROUP	N	VE Mean	VE SD
1	1	50	4.16000000	3.46622184
1	2	40	4.72500000	3.88941694
2	1	50	4.50000000	3.97055489
2	2	40	5.20000000	3.89081759

Level of ADMIN	Level of MODE	Level of GROUP	N	VE Mean	VE SD
1	1	1	50	4.16000000	3.46622184
1	2	2	40	5.20000000	3.89081759
2	1	2	40	4.72500000	3.88941694
2	2	1	50	4.50000000	3.97055489

TABLE 32

SUMMARY OF MEANS
VPI - CONVENTIONAL SCALE SCORES

Level of ADMIN	N	VC Mean	VC SD
1	90	3.56666667	3.70862948
2	90	3.63333333	4.01807713

Level of MODE	N	VC Mean	VC SD
1	90	3.33333333	3.66274901
2	90	3.86666667	4.04247115

Level of GROUP	N	VC Mean	VC SD
1	100	3.77000000	4.03983948
2	80	3.38750000	3.62653285

Level of ADMIN	Level of MODE	N	VC Mean	VC SD
1	1	50	3.50000000	3.75390953
1	2	40	3.65000000	3.69719228
2	1	40	3.12500000	3.58191683
2	2	50	4.04000000	4.32817066

Level of ADMIN	Level of GROUP	N	VC Mean	VC SD
1	1	50	3.50000000	3.75390953
1	2	40	3.65000000	3.69719228
2	1	50	4.04000000	4.32817066
2	2	40	3.12500000	3.58191683

Level of MODE	Level of GROUP	N	VC Mean	VC SD
1	1	50	3.50000000	3.75390953
1	2	40	3.12500000	3.58191683
2	1	50	4.04000000	4.32817066
2	2	40	3.65000000	3.69719228

Level of ADMIN	Level of MODE	Level of GROUP	N	VC Mean	VC SD
1	1	1	50	3.50000000	3.75390953
1	2	2	40	3.65000000	3.69719228
2	1	2	40	3.12500000	3.58191683
2	2	1	50	4.04000000	4.32817066

TABLE 33

SUMMARY OF MEANS
VPI - SELF-CONTROL SCALE SCORES

Level of ADMIN	N	VSC Mean	VSC SD
1	90	8.76666667	3.83962194
2	90	8.56666667	3.83083292

Level of MODE	N	VSC Mean	VSC SD
1	90	8.82222222	3.67064816
2	90	8.51111111	3.98943674

Level of GROUP	N	VSC Mean	VSC SD
1	100	8.37000000	3.96106046
2	80	9.03750000	3.64029401

Level of ADMIN	Level of MODE	N	VSC Mean	VSC SD
1	1	50	8.60000000	3.85449645
1	2	40	8.97500000	3.85963662
2	1	40	9.10000000	3.45520789
2	2	50	8.14000000	4.09085591

Level of ADMIN	Level of GROUP	N	VSC Mean	VSC SD
1	1	50	8.60000000	3.85449645
1	2	40	8.97500000	3.85963662
2	1	50	8.14000000	4.09085591
2	2	40	9.10000000	3.45520789

Level of MODE	Level of GROUP	N	VSC Mean	VSC SD
1	1	50	8.60000000	3.85449645
1	2	40	9.10000000	3.45520789
2	1	50	8.14000000	4.09085591
2	2	40	8.97500000	3.85963662

Level of ADMIN	Level of MODE	Level of GROUP	N	VSC Mean	VSC SD
1	1	1	50	8.60000000	3.85449645
1	2	2	40	8.97500000	3.85963662
2	1	2	40	9.10000000	3.45520789
2	2	1	50	8.14000000	4.09085591

TABLE 34

SUMMARY OF MEANS
VPI - MASCULINITY-FEMININITY SCALE SCORES

Level of ADMIN	N	VMF Mean	VMF SD
1	90	9.00000000	2.28330531
2	90	9.08888889	2.04249486

Level of MODE	N	VMF Mean	VMF SD
1	90	9.06666667	2.28232092
2	90	9.02222222	2.04432774

Level of GROUP	N	VMF Mean	VMF SD
1	100	8.90000000	2.16258357
2	80	9.22500000	2.15814618

Level of ADMIN	Level of MODE	N	VMF Mean	VMF SD
1	1	50	8.88000000	2.37881809
1	2	40	9.15000000	2.17856692
2	1	40	9.30000000	2.16261949
2	2	50	8.92000000	1.94663498

Level of ADMIN	Level of GROUP	N	VMF Mean	VMF SD
1	1	50	8.88000000	2.37881809
1	2	40	9.15000000	2.17856692
2	1	50	8.92000000	1.94663498
2	2	40	9.30000000	2.16261949

Level of MODE	Level of GROUP	N	VMF Mean	VMF SD
1	1	50	8.88000000	2.37881809
1	2	40	9.30000000	2.16261949
2	1	50	8.92000000	1.94663498
2	2	40	9.15000000	2.17856692

Level of ADMIN	Level of MODE	Level of GROUP	N	VMF Mean	VMF SD
1	1	1	50	8.88000000	2.37881809
1	2	2	40	9.15000000	2.17856692
2	1	2	40	9.30000000	2.16261949
2	2	1	50	8.92000000	1.94663498

TABLE 35

SUMMARY OF MEANS
VPI - STATUS SCALE SCORES

Level of ADMIN	N	VST Mean	VST SD
1	90	9.03333333	2.40060854
2	90	9.24444444	2.33349384

Level of MODE	N	VST Mean	VST SD
1	90	9.13333333	2.34257133
2	90	9.14444444	2.39644452

Level of GROUP	N	VST Mean	VST SD
1	100	9.14000000	2.64773583
2	80	9.13750000	1.96645121

Level of ADMIN	Level of MODE	N	VST Mean	VST SD
1	1	50	9.04000000	2.66465911
1	2	40	9.02500000	2.05672755
2	1	40	9.25000000	1.89127552
2	2	50	9.24000000	2.65391509

Level of ADMIN	Level of GROUP	N	VST Mean	VST SD
1	1	50	9.04000000	2.66465911
1	2	40	9.02500000	2.05672755
2	1	50	9.24000000	2.65391509
2	2	40	9.25000000	1.89127552

Level of MODE	Level of GROUP	N	VST Mean	VST SD
1	1	50	9.04000000	2.66465911
1	2	40	9.25000000	1.89127552
2	1	50	9.24000000	2.65391509
2	2	40	9.02500000	2.05672755

Level of ADMIN	Level of MODE	Level of GROUP	N	VST Mean	VST SD
1	1	1	50	9.04000000	2.66465911
1	2	2	40	9.02500000	2.05672755
2	1	2	40	9.25000000	1.89127552
2	2	1	50	9.24000000	2.65391509

TABLE 36

SUMMARY OF MEANS
VPI - INFREQUENCY SCALE SCORES

Level of ADMIN	N	VINF Mean	VINF SD
1	90	5.65555556	2.38987147
2	90	5.54444444	2.27917347

Level of MODE	N	VINF Mean	VINF SD
1	90	5.54444444	2.40393456
2	90	5.65555556	2.26433566

Level of GROUP	N	VINF Mean	VINF SD
1	100	5.32000000	2.19678369
2	80	5.95000000	2.45413627

Level of ADMIN	Level of MODE	N	VINF Mean	VINF SD
1	1	50	5.32000000	2.32501920
1	2	40	6.07500000	2.43255168
2	1	40	5.82500000	2.50012820
2	2	50	5.32000000	2.08434397

Level of ADMIN	Level of GROUP	N	VINF Mean	VINF SD
1	1	50	5.32000000	2.32501920
1	2	40	6.07500000	2.43255168
2	1	50	5.32000000	2.08434397
2	2	40	5.82500000	2.50012820

Level of MODE	Level of GROUP	N	VINF Mean	VINF SD
1	1	50	5.32000000	2.32501920
1	2	40	5.82500000	2.50012820
2	1	50	5.32000000	2.08434397
2	2	40	6.07500000	2.43255168

Level of ADMIN	Level of MODE	Level of GROUP	N	VINF Mean	VINF SD
1	1	1	50	5.32000000	2.32501920
1	2	2	40	6.07500000	2.43255168
2	1	2	40	5.82500000	2.50012820
2	2	1	50	5.32000000	2.08434397

TABLE 37

SUMMARY OF MEANS
VPI - ACQUIESCENCE SCALE SCORES

Level of ADMIN	N	VAC Mean	VAC SD
1	90	11.5111111	5.11715433
2	90	11.6444444	5.29852876

Level of MODE	N	VAC Mean	VAC SD
1	90	11.0222222	5.09014862
2	90	12.1333333	5.26638695

Level of GROUP	N	VAC Mean	VAC SD
1	100	12.1800000	5.36803107
2	80	10.8250000	4.89839808

Level of ADMIN	Level of MODE	N	VAC Mean	VAC SD
1	1	50	11.6200000	5.27175764
1	2	40	11.3750000	4.98041034
2	1	40	10.2750000	4.81444142
2	2	50	12.7400000	5.45785336

Level of ADMIN	Level of GROUP	N	VAC Mean	VAC SD
1	1	50	11.6200000	5.27175764
1	2	40	11.3750000	4.98041034
2	1	50	12.7400000	5.45785336
2	2	40	10.2750000	4.81444142

Level of MODE	Level of GROUP	N	VAC Mean	VAC SD
1	1	50	11.6200000	5.27175764
1	2	40	10.2750000	4.81444142
2	1	50	12.7400000	5.45785336
2	2	40	11.3750000	4.98041034

Level of ADMIN	Level of MODE	Level of GROUP	N	VAC Mean	VAC SD
1	1	1	50	11.6200000	5.27175764
1	2	2	40	11.3750000	4.98041034
2	1	2	40	10.2750000	4.81444142
2	2	1	50	12.7400000	5.45785336

References

Allred, Linda J. *Sources of Non-equivalence Between Computerized and Conventional Psychological Tests*. PhD dissertation. The John Hopkins University, Baltimore MD, 1986 (UMI-8615946).

Anastasi, Anne. *Psychological Testing* (Fifth Edition). New York: McMillan Publishing Co., Inc., 1982.

Bagley, Christopher and Mark Genuis. "Psychology of Computer Use: XX. Sexual Abuse Recalled: Evaluation of a Computerized Questionnaire in a Population of Young Adult Males," *Perceptual and Motor Skills, 72*: 287-288 (February 1991).

Brown, Douglas T. "Automated Assessment Systems in School and Clinical Psychology: Present Status and Future Directions," *School Psychology Review, 13*: 455-460 (Fall 1984).

Burke, Michael J. and Jacques Normand. "Computerized Psychological Testing: Overview and Critique," *Professional Psychology: Research and Practice, 18*: 42-51 (February 1987).

Campbell, Donald T. and Julian C. Stanley. *Experimental and Quasi-experimental Designs for Research.* Boston: Houghton Mifflin Company, 1963.

Casey, Major Maureen C. and Captain Douglas A. Kveene. *A Comparative Analysis of Computer End-User Support in the Air Force and Civilian Organizations*. MS Thesis, AFIT/GIR/LSR/91D-2. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1991 (AD-A244118).

Elithorn, Alick, Sue Mornington, and Andreas Stavrou. "Automated Psychological Testing: Some Principles and Practice," *Journal of Man-Machine Studies, 17*: 247-263 (October 1982).

Emory, C. William and Donald R. Cooper. *Business Research Methods* (Fourth Edition). Homewood IL: Irwin, 1991.

Farrell, Albert D. "Impact of Standards for Computer-Based Tests on Practice: Consequences of the Information Gap," *Computers in Human Behavior, 5*: 1-11 (1989).

Guastello, Stephen J., Denise D. Guastello, and Larry L. Craft. "Assessment of the Barnum Effect in Computer-Based Test Interpretations," *Journal of Psychology, 71*: 477-484 (September 1989).

Hedlund, James L. "Mental Health Computing: Directions for Research," *Computers in Human Services, 2*: 9-36 (Fall-Winter 1987).

-----. "Mental Health Computing in Great Britain," *Computers in Human Services, 3*: 5-27 (1988).

Holland, John L. *Vocational Preference Inventory (VPI): Professional Manual* (1985 Edition). Odessa FL: Psychological Assessment Resources, Inc., 1985.

Hopkins, Kenneth D. and Julian C. Stanley. *Educational and Psychological Measurement and Evaluation* (Sixth Edition). Englewood Cliffs NJ: Prentice-Hall, Inc., 1981.

Kennedy, Robert S., Dennis R. Baltzley, Janet J. Turnage, and Marshall B. Jones. "Factor Analysis and Predictive Validity of Microcomputer-based Tests," *Perceptual and Motor Skills, 69*: 1059-1074 (December 1989a, part 2).

-----, Dennis R. Baltzley, Robert L. Wilkes, and Lois A. Kuntz. "Psychology of Computer Use: IX. A Menu of Self-Administered Microcomputer-Based Neurotoxicology Tests," *Perceptual and Motor Skills, 68*: 1255-1272 (June 1989b, part 2).

Kramer, Jack J. "On the Question of Professional Standards for Computer-Based Test Interpretation," *American Psychologist, 42*: 889-890 (September 1987).

Lautenschlager, Gary J. and Vicki L. Flaherty. "Computer Administration of Questions: More Desirable or More Social Desirability," *Journal of Applied Psychology, 75*: 310-314 (June 1990).

Matarazzo, Joseph D. "Computerized Clinical Psychological Test Interpretations: Unvalidated Plus All Mean and No Sigma," *American Psychologist, 41*: 14-24 (January 1986).

McCaughan, Robert A. *Computerized Testing Impact on Expectancy.* PhD dissertation. University of Denver, Denver CO, 1982 (UMI-8216692).

McClave, James T. and P. George Benson. *Statistics for Business and Economics* (Fifth Edition). San Francisco: Dellen Publishing Company, 1991.

Schinka, John A. *Vocational Preference Inventory: Computer Version Manual.* IBM Version 3.0 Manual A 050988, Odessa FL: Psychological Assessment Resources, Inc., 1989.

Space, Lawrence G. "The Computer as Psychometrician," *Behavioral Research Methods & Instrumentation, 13*: 595-606 (August 1981).

Thompson, James A. and Sarah L. Wilson. "Automated Psychological Testing," *International Journal of Man-Machine Studies, 17*: 279-289 (October 1982).

Townsend, John C. *Introduction to Experimental Method: For Psychology and the Social Sciences*. New York: McGraw-Hill Book Company, Inc., 1953.

Vale, C. David. "Design and Implementation of a Microcomputer-Based Adaptive Testing System," *Behavioral Research Methods & Instrumentation, 13*: 399-406 (August 1981).

Volans, Pamela Jane. "Pros and Cons of Tailored Testing: An Examination of Issues Highlighted by Experience with an Automated Testing System," *International Journal of Man-Machine Studies, 17*: 301-304 (October 1982).

Walker, N. William and Carolyn Cobb Myrick. "Ethical Considerations in the Use of Computers in Psychological Testing and Assessment," *Journal of School Psychology, 23*: 51-57 (Spring 1985).

Waller, Niels G. and Steve P. Reise. "Computerized Adaptive Personality Assessment: An Illustration With the Absorption Scale," *Journal of Personality and Social Psychology, 57*: 1051-1058 (December 1989).

Wilson, Sarah L., James A. Thompson, and Geoffrey Wylie. "Automated Psychological Testing for the Severely Physically Handicapped," *International Journal of Man-Machine Studies, 17*: 291-296 (October 1982).

Wonderlic, E.F. *Wonderlic Personnel Test Manual*. Northfield IL: E.F. Wonderlic & Associates, Inc., 1983.

Wonderlic Personnel Test, Inc. *The Wonderlic Personnel Test -- PC Version Manual*. Personal Computer Version 2.2 (WPT-PC). Northfield IL: Wonderlic Personnel Test, Inc., 1992.

Vita

Captain Timothy R. Morris was born in Fairfax, Virginia on October 21, 1964. After graduating from Carterville High School, Carterville, Illinois, in 1982, he attended Southern Illinois University at Carbondale, Illinois. He graduated in August 1986, with a Bachelor of Arts Degree in Sociology. After completion of Officers Training School at Lackland Air Force Base, Texas, in November of 1987, he was commissioned a Second Lieutenant. In April of 1988, he was assigned to Los Angeles Air Force Base, California, Headquarters Space Systems Division. From April 1988 to October 1988 he worked as the Assistant Project Manager for the Delta II Space Launch Vehicle Program. In October 1988 he became the Executive Officer for the Medium Launch Vehicles System Program Office. In April of 1989 he was named the Project Manager, Atlas E Space Launch Vehicle Program, Atlas Launch Vehicle Systems Program Office. In May 1991, Lieutenant Morris entered the Air Force Institute of Technology as a master's candidate in the Logistics Management Program of the School of Systems and Logistics. In November 1991, Lieutenant Morris was promoted to the rank of Captain. In December of 1991, Captain Morris was inducted into the National Honorary Fraternity for Management, Sigma Iota Epsilon. In April 1992, he was selected for honorary recognition of academic achievement in *THE NATIONAL DEAN'S LIST*, 1991-1992.

Permanent Address: R.R. 2 Box 214
Carterville, IL 62918

Vita

Captain Anthony C. Hensley was born in Burlington, North Carolina on August 11, 1964. After graduating from Western Alamance High School, Elon College, North Carolina, in 1982, he attended East Carolina University in Greenville, North Carolina. He graduated in May 1986 with a Bachelor of Science Degree in Political Science. Upon graduation, he was commissioned a Second Lieutenant. In January of 1987 he was assigned to the 31st Tactical Fighter Wing, Homestead Air Force Base, Florida. From January 1987 to February 1988 he worked as Officer-In-Charge 31st Combat Support Unit. In February 1988 he became Assistant Officer-In-Charge, 309th Aircraft Maintenance Unit. In May of 1989 he was assigned to the Deputy Commander For Maintenance staff as Officer-In-Charge, Maintenance Operations Center. In May 1990 he became Officer-In-Charge of the 31st Propulsion Branch, during which time he was chosen to manage the 31st Wing's maintenance conversion to Block 40 F-16 aircraft. In May 1991, Captain Hensley entered the Air Force Institute of Technology as a master's candidate in the Logistics Management Program of the School of Systems and Logistics. In December of 1991, Captain Hensley was inducted into the National Honorary Fraternity for Management, Sigma Iota Epsilon.

Permanent Address: 712 Greenwood Drive
Burlington, NC 27217

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1992	Master's Thesis

4. TITLE AND SUBTITLE

A COMPARISON OF COMPUTER AND PAPER MEDIA IN PSYCHOLOGICAL TESTING

5. FUNDING NUMBERS

6. AUTHOR(S)

Anthony C. Hensley, Captain, USAF
Timothy R. Morris, Captain, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology, WPAFB OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GLM/LSR/92S-22

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This study examined differences based on computer and paper modes of administration in psychological testing. The hypotheses tested were the expectations that scores on the computer Wonderlic Personnel Test (WPT) would be higher than the paper version and subjects would report higher levels on the computer Vocational Preference Inventory (VPI) scales. A counterbalance experimental design with two experimental groups was used. The subjects were volunteers from the Air Force Institute of Technology (AFIT) and Wright State University (WSU); 90 subjects completed the study. This included 75 students from AFIT and 15 from WSU; 69 subjects were male and 21 were female. Subjects were randomly assigned to two experimental groups. The administration mode of the tests was the independent variable for the study. It was concluded that there was no significant difference between the modes of administration for either the WPT or the VPI. There was a significant order effect with the WPT. Subjects scored significantly higher on the WPT retest. There were also significant differences in four VPI response variables between experimental groups. Subjects preferred the computer administration over the paper administration 54% to 17%. 50% reported the computer versions "easier to complete", while 13% reported the paper versions "easier to complete".

14. SUBJECT TERMS	15. NUMBER OF PAGES
Psychology, Psychological Testing, Computer, Computer Testing, Psychometrics, Human Factors	112
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

AFIT Control Number AFIT/GLM/LSR/92S-22

AFIT RESEARCH ASSESSMENT

The purpose of this questionnaire is to determine the potential for current and future applications of AFIT thesis research. Please return completed questionnaires to: AFIT/LSC, Wright-Patterson AFB OH 45433-9905.

1. Did this research contribute to a current research project?

a. Yes b. No

2. Do you believe this research topic is significant enough that it would have been researched (or contracted) by your organization or another agency if AFIT had not researched it?

a. Yes b. No

3. The benefits of AFIT research can often be expressed by the equivalent value that your agency received by virtue of AFIT performing the research. Please estimate what this research would have cost in terms of manpower and/or dollars if it had been accomplished under contract or if it had been done in-house.

Man Years _____ $ _____

4. Often it is not possible to attach equivalent dollar values to research, although the results of the research may, in fact, be important. Whether or not you were able to establish an equivalent value for this research (3, above) what is your estimate of its significance?

a. Highly Significant b. Significant c. Slightly Significant d. Of No Significance

5. Comments

_____ _____

Name and Grade Organization

_____ _____

Position or Title Address